MAXWELL BENNETT, DANIEL DENNETT
PETER HACKER Y JOHN SEARLE

LA NATURALEZA DE LA CONCIENCIA

Cerebro, mente y lenguaje

PAIDÓS
Barcelona
Buenos Aires
México

PAIDÓS TRANSICIONES

Últimos títulos publicados

29. H. Gardner, *La inteligencia reformulada*
30. J. Horgan, *La mente por descubrir*
31. R. Lewontin, *El sueño del genoma humano y otras ilusiones*
32. J. Elster, *Sobre las pasiones*
33. U. Bronfenbrenner, *La ecología del desarrollo humano*
34. D. A. Norman, *El ordenador invisible*
35. K. Davies, *La conquista del genoma humano*
36. G. Claxton, *Aprender*
37. C. Hagège, *No a la muerte de las lenguas*
38. C. Korsmeyer, *El sentido del gusto*
39. E. S. Hochman, *La Bauhaus. Crisol de la modernidad*
40. M. Rees, *Nuestro hábitat cósmico*
41. H. Gardner y otros, *Buen trabajo*
42. F. de Waal, *El simio y el aprendiz de sushi*
43. R. M. Hogarth, *Educar la intuición*
44. S. Weinberg, *Plantar cara*
45. A. C. Danto, *La Madonna del futuro*
46. N. Longworth, *El aprendizaje a lo largo de la vida*
48. M. Levine, *Mentes diferentes, aprendizajes diferentes*
49. S. Pinker, *La tabla rasa*
50. R. Aunger, *El meme eléctrico*
51. D. C. Dennett, *La evolución de la libertad*
52. M. Levine, *Contra el mito de la pereza*
53. R. Debray y J. Bricmont, *A la sombra de la ilustración*
54. H. Gardner, *Mentes flexibles*
55. G. Nunberg (comp.), *El futuro del libro*
56. N. Longworth, *El aprendizaje a lo largo de la vida en la práctica*
57. C. Allègre, *Un poco de ciencia para todo el mundo*
58. D. A. Norman, *El diseño emocional*
59. D. J. Watts, *Seis grados de separación. La ciencia de las redes en la era del acceso*
60. M. P. Lynch, *La importancia de la verdad*
61. M. S. Gazzaniga, *El cerebro ético*
62. H. Gee, *La escalera de Jacob. Historia del genoma humano*
63. G. Rizzolatti y C. Sinigaglia, *Las neuronas espejo. Los mecanismos de la empatía emocional*
64. R. Sapolsky, *El mono enamorado y otros ensayos sobre nuestra vida animal*
65. C. Allègre, *La sociedad vulnerable. Doce retos de política científica*
66. F. de Waal, *Primates y filósofos. La evolución de la moral del simio al hombre*
67. S. Pinker, *El mundo de las palabras. Una introducción a la naturaleza humana*
68. J. Dewey, *Cómo pensamos*
69. N. N. Taleb, *El cisne negro*
70. B. R. Komisaruk y otros, *La ciencia del orgasmo*
71. H. Gardner, *Las cinco mentes del futuro*
72. M. Bennett y otros, *La naturaleza de la conciencia*

Título original: *Neuroscience & Philosophy. Brain, Mind & Language*
Publicado originalmente en inglés, en 2007, por Columbia University Press, Nueva York

Traducción de Roc Filella
Revisión técnica de Francesc Forn

Cubierta de excèntric comunicació

Quedan rigurosamente prohibidas, sin la autorización escrita de los titulares del «Copyright», bajo las sanciones establecidas en las leyes, la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la reprografía y el tratamiento informático, y la distribución de ejemplares de ella mediante alquiler o préstamo públicos.

© 2003 de los extractos de *Philosophical Foundations of Neuroscience* by Maxwell R. Bennett and Peter M. S. Hacker
© 2007 de «Philosophy as Naive Anthropology: Comment on Bennett and Hacker» by Daniel Dennett
© 2007 de «Putting Conclousness Back in the Brain: Reply to Bennett and Hacker, *Philosophical Foundations of Neuroscience*» by John Searle
© 2007 de «The Conceptuals Presuppositions of Cognitive Neuroscience: A Reply to Critics» by Maxwell R. Bennett and Peter M. S. Hacker
© 2007 del epílogo by Maxwell R. Bennett
© 2007 de la «Introduction» y «Still Looking: Science and Philosophy in Pursuit of Prince Reason» by Columbia University Press. All rights reserved
© 2008 de la traducción, Roc Filella
© 2008 de todas las ediciones en castellano,
Ediciones Paidós Ibérica, S.A.,
Av. Diagonal, 662-664 - 08034 Barcelona
www.paidos.com

ISBN: 978-84-493-2132-0
Depósito legal: B-18.095/2008

Impreso en Cayfosa-Quebecor, S.A.
Crta. de Caldes, km. 3,7 - 08130 Santa Perpètua de Mogoda (Barcelona)

Impreso en España – Printed in Spain

SUMARIO

Introducción 9
Daniel Robinson

LA POLÉMICA

Fragmentos de *Philosophical Foundations of Neuroscience* ... 15
Maxwell Bennett y Peter Hacker
Neurociencia y filosofía 69
Maxwell Bennett

LAS REFUTACIONES

La filosofía como antropología ingenua 93
Daniel Dennett
Situar de nuevo la conciencia en el cerebro 121
John Searle

RÉPLICA A LAS REFUTACIONES

Los supuestos conceptuales de la neurociencia cognitiva ... 159
Maxwell Bennett y Peter Hacker

Epílogo 203
MAXWELL BENNETT

La búsqueda continúa 215
DANIEL ROBINSON

Notas 245

INTRODUCCIÓN

DANIEL ROBINSON

La editorial Blackwell publicó, en 2003, *Philosophical Foundations of Neuroscience*, de Max Bennett y Peter Hacker. El libro llamó la atención de inmediato, porque era la primera evaluación sistemática de las bases conceptuales de la neurociencia, tal como habían sido establecidas por científicos y filósofos. Lo que añadía atractivo a la obra eran dos apéndices dedicados a la crítica en detalle de los influyentes escritos de John Searle y Daniel Dennett. Max Bennett, consumado neurocientífico, identificó con acierto a Searle y Dennett como los filósofos más leídos entre la comunidad neurocientífica, y tenía mucho interés en dejar claro a los lectores por qué él y Hacker disentían de sus opiniones.

En el otoño de 2004, la comisión de programación de la American Philosophical Association (APA) invitó a Bennett y a Hacker a participar en una sesión de «Autores y críticos» en el encuentro de 2005 de la Asociación, en Nueva York. La elección de los críticos no pudo haber sido mejor: Daniel Dennett y John Searle aceptaron escribir sus réplicas a las críticas que Bennett y Hacker habían formulado contra su obra. El contenido de este libro se basa en aquella sesión de tres horas de la APA. Owen Flanagan presidió la sesión, que estuvo marcada por un intercambio inusualmente animado entre los participantes. Antes de la sesión Dennett y Searle habían facilitado sendas versiones escritas de sus refutaciones, a las que Bennett y Hacker replicaron a continuación.

Plenamente consciente de la importancia de los temas filosóficos debatidos, Wendy Lochner, editora de la sección de filosofía de Columbia University Press, propuso a los participantes que las actas se publicaran en forma de libro. Por lo general, un debate en origen apasionado suele perder colorido y emoción en su versión escrita. El lector se ve forzado a recrear en su imaginación un acontecimiento real a partir de la fragmentación y el artificio de la prosa editada. Creo que es justo decir que este libro no padece de esta limitación tan común. En los presentes ensayos e intercambios, el lector reconocerá el poder motivador de la pasión intelectual. Los participantes abordan sus temas en profundidad. Sus destacadas aportaciones a lo largo de varias décadas les dan derecho a que se les tome en serio. Además, hay mucho más en juego de lo que es habitual en este tipo de discusiones. Al fin y al cabo, el proyecto de la neurociencia cognitiva es nada menos que la incorporación a la estructura de la propia ciencia de lo que nos complacemos en llamar naturaleza humana. Dennett y Searle, con una confianza que puede parecer entusiasmo, se inclinan a creer que el proceso de incorporación va por buen camino. Por su parte Bennett y Hacker, con una cautela que puede parecer escepticismo, plantean la posibilidad de que el proyecto mismo esté basado en una confusión.

Fui honrado con la petición de redactar un capítulo final para el libro proyectado. En él resumo mis concepciones más arraigadas sobre el tema, al tiempo que sopeso los diestros argumentos y contraargumentos de las principales figuras que intervienen en el debate. Confío en que los lectores se darán cuenta de que tengo pocas ideas definitivas sobre el particular. Reconozco el decidido empeño de Searle y Dennett por ofrecer un modelo creíble y viable de cómo la vida mental se comprende gracias a acontecimientos que se producen debajo de la piel. Norbert Wiener, uno de los hombres de cien-

cia sabio de verdad, señaló que el mejor modelo material de un gato es un gato, a ser posible el mismo gato. No obstante, sin modelos —incluso aquellos aderezados con tintes antropomórficos—, la misma complejidad del mundo real acaba frustrando el progreso científico en cualquier campo. No existen cálculos ni ecuaciones que establezcan las fronteras en cuyo interior deba confinarse la imaginación de quien construye el modelo.

Al final, este tipo de cuestiones se zanjan en el ámbito de la estética. Con ello no quiero decir que haya menos espacio para el rigor analítico: en su máxima expresión, el análisis filosófico es una tarea estética. Seguramente, esto es lo que hace que el físico y el matemático se sientan atraídos por lo «elegante». ¿No es la estética la que emplea la navaja de Occam para decidir sobre el refinamiento, la medida, la proporcionalidad y la coherencia? Estoy convencido de que, justamente bajo estos criterios, los lectores encontrarán en la crítica de Bennett-Hacker —en particular en la filosóficamente rica e informada crítica de Peter Hacker— no una inclinación al escepticismo, sino una aplicación cuidadosa y efectivamente *elegante* de las mejores herramientas ideadas por los filósofos.

Dicho esto, es importante ir más allá y reconocer que no es probable que nuestra vida realmente *vivida* desvele su realidad completa, cambiante, a menudo voluble y maravillosamente interior, bien sea a las tablas veritativas, a la máquina de Turing o al laparoscopio. No es de extrañar que el filósofo que tan a menudo nos ha proporcionado la primera reflexión sobre un tema importante sea también quien tenga la última palabra sobre el mismo. Me refiero, por supuesto, a Aristóteles. Debemos buscar la precisión en aquellas cosas que la admiten, y escoger las herramientas adecuadas para la tarea que tenemos entre manos. En última instancia, nuestras explicaciones deben mantener una relación inteligible con aquello que pretende-

mos explicar. El demógrafo que afirma con una precisión encomiable que la familia media se compone de 2,53 miembros no cree necesario recordarnos que no hay tal cosa como un 0,53 de persona. Con este tipo de datos no se pretende describir la naturaleza de los elementos contados; esa cantidad es simplemente su resultado. La cuestión es, claro está, que la precisión científica o, para el caso, la precisión aritmética, no pueden decirnos casi nada sobre qué es lo que se ha analizado con tal precisión. Aquí como en cualquier otra parte, la máxima que rige es la de *caveat emptor* [tenga cuidado el comprador].

Los lectores abordarán este discurso con auténtico interés, incluso con un poco de vanidad, ya que trata de *ellos*. Son sus propios criterios estéticos los que se impondrán a un material de este tipo, y son ellos quienes finalmente decidirán si las explicaciones ofrecidas guardan una relación inteligible con lo que realmente importa. Pero un buen jurado debe atenerse a las pruebas disponibles, dejarse guiar en sus deliberaciones por el sólido criterio de la evidencia. ¡Paciente lector! ¡Digno miembro del jurado! He aquí algunas pruebas (procedentes de la neurociencia cognitiva) y una exposición excepcionalmente clara de las reglas que cabe aplicar en su consideración. No es necesario darse prisa en emitir un veredicto...

DANIEL N. ROBINSON

LA POLÉMICA

PHILOSOPHICAL FOUNDATIONS OF NEUROSCIENCE

Introducción

MAXWELL BENNETT Y PETER HACKER

El libro *Philosophical Foundations of Neuroscience* [Fundamentos filosóficos de la neurociencia] expone los frutos de un trabajo llevado a cabo conjuntamente por un neurocientífico y un filósofo. Trata de los fundamentos conceptuales de la neurociencia *cognitiva*, unos fundamentos constituidos por las relaciones estructurales entre los conceptos psicológicos que intervienen en las investigaciones sobre las bases neurales de las capacidades cognitivas, afectivas y volitivas humanas. Investigar las relaciones lógicas entre los conceptos es una tarea filosófica. Orientar esta investigación por unos determinados senderos, de modo que se vaya a esclarecer la investigación del cerebro es una tarea neurocientífica. De ahí nuestra empresa conjunta.

Si queremos comprender las estructuras y la dinámica neurales que hacen posible la percepción, el pensamiento, la memoria, el sentimiento y la conducta intencional, es esencial la claridad sobre estos conceptos y categorías. Ambos autores, que abordaban esta investigación desde perspectivas muy diferentes, se sintieron desconcertados, y a veces incómodos, con el uso que de los conceptos psicológicos se hace en la neurociencia actual. El desconcierto lo provocaban las implicaciones de lo que cierto neurocientífico afirmaba sobre el cerebro y la mente, o las razones por las que un neurocientífico creía

que sus experimentos esclarecían la capacidad psicológica objeto de estudio, o los presupuestos conceptuales de las preguntas que se formulaban. La incomodidad provenía de la sospecha de que, en algunos casos, los conceptos no se entendían, se aplicaban incorrectamente o más allá de su campo específico de aplicación. Y cuanto más investigábamos, más convencidos estábamos de que, pese a los impresionantes avances de la neurociencia cognitiva, algo andaba mal en la teorización general.

Las preguntas empíricas sobre el sistema nervioso son el campo de la neurociencia. El cometido de ésta es establecer los hechos en lo que concierne a las estructuras y las operaciones neurales. Es tarea de la neurociencia *cognitiva* explicar las condiciones neurales que hacen posibles las funciones perceptivas, cognitivas, cogitativas, afectivas y volitivas. Las investigaciones experimentales confirman o cuestionan tales teorías explicativas. Por contraste, las preguntas conceptuales (las que, por ejemplo, se refieren a los conceptos de mente o memoria, pensamiento o imaginación), la descripción de las relaciones lógicas entre los conceptos (como las que existen entre los conceptos de percepción y sensación, o los de conciencia y autoconciencia) y el examen de las relaciones estructurales entre los distintos campos conceptuales (por ejemplo, entre el psicológico y el neural, o el mental y el conductista), son el campo propio de la filosofía.

Las preguntas conceptuales son previas a las cuestiones de verdad y falsedad. Son preguntas que conciernen a nuestras *formas de representación*, no a la verdad o la falsedad de afirmaciones empíricas. Estas formas están presupuestas en las afirmaciones científicas verdaderas (*y* en las falsas) y las teorías científicas correctas (*e* incorrectas). No determinan lo que es empíricamente verdadero o falso, sino más bien lo que tiene y lo que no tiene sentido. De ahí que las preguntas conceptuales no sean pertinentes en la investigación y la ex-

perimentación científicas ni en la teorización científica. Y es que cualquiera de estas investigaciones y teorizaciones *presuponen* los conceptos y las relaciones conceptuales en cuestión. Nuestro interés aquí no está en las líneas de demarcación gremiales, sino en las distinciones entre tipos lógicamente diferentes de indagación intelectual.[1]

Es de suma importancia distinguir las preguntas conceptuales de las empíricas. Cuando una pregunta conceptual se confunde con una científica, es inevitable que parezca singularmente abstrusa. En estos casos, parece como si la ciencia debiera ser capaz de descubrir la verdad de lo que se investiga mediante la teoría y el experimento, y sin embargo fracasa continuamente en tal empeño. Y no hay que extrañarse de ello, ya que las preguntas conceptuales no pueden tratarse con los métodos empíricos de investigación, del mismo modo que los problemas de las matemáticas puras no son resolubles por los métodos de la física. Además, cuando los problemas empíricos se abordan sin la adecuada claridad conceptual, forzosamente se plantean preguntas equívocas que conducen a la investigación por caminos infructuosos. Cualquier oscuridad en lo que respecta a los conceptos relevantes se reflejará en una correspondiente falta de claridad en las preguntas y, por lo tanto, en el diseño de los experimentos con los que se pretende responder a esas preguntas. Y cabe prever que cualquier incoherencia en la comprensión de la estructura conceptual relevante se manifieste en incoherencias en la interpretación de los resultados de los experimentos.

La neurociencia cognitiva opera a lo ancho de la frontera entre dos campos, la neurofisiología y la psicología, cuyos respectivos conceptos son categóricamente disímiles. Las relaciones lógicas o conceptuales entre lo fisiológico y lo psicológico son problemáticas. Numerosos conceptos y categorías de conceptos psicológicos son difíciles de definir con nitidez. Las relaciones entre la mente y el cerebro, y en-

tre lo psicológico y lo conductual, son desconcertantes. La extrañeza ante estos conceptos y sus expresiones, y ante estos aparentes «dominios» y sus relaciones, ha sido una característica de la neurofisiología desde sus orígenes.[2] A pesar de los grandes avances de la neurociencia a principios del siglo XX de la mano de Charles Sherrington, la batería de preguntas conceptuales popularmente conocidas como el problema mente-cuerpo o mente-cerebro siguió tan inabordable como siempre, como se pone de manifiesto en las erradas ideas cartesianas defendidas por Sherrington y algunos de sus colegas y protegidos, como Edgar Adrian, John Eccles y Wilder Penfield. Pese a la incuestionable brillantez de su obra, permanecían graves confusiones conceptuales.[3] El objeto de nuestra investigación en este libro es determinar si la generación de neurocientíficos actual ha superado con éxito las confusiones conceptuales de generaciones anteriores, o se ha limitado a sustituir una mezcolanza conceptual por otras.

Tal mezcolanza es evidente en la persistente adscripción de atributos psicológicos al cerebro. Mientras Sherrington y sus protegidos adscribían atributos psicológicos a la mente (concebida como una sustancia peculiar, tal vez inmaterial, distinta del cerebro), los neurocientíficos actuales tienden a adscribir la misma serie de atributos psicológicos al cerebro (del que se piensa comúnmente, aunque no de forma uniforme, que es idéntico a la mente). Pero la mente, sostenemos nosotros,[4] ni es una sustancia distinta del cerebro ni una sustancia idéntica al cerebro. Y demostramos que la adscripción de atributos psicológicos al cerebro es incoherente.[5] Los seres humanos poseemos una amplia diversidad de capacidades psicológicas, que se ejercen en las circunstancias de la vida, cuando percibimos, pensamos y razonamos, sentimos emociones, deseamos cosas, hacemos planes y tomamos decisiones. La posesión y el ejercicio de esas ca-

pacidades nos definen como el tipo de animal que somos. Podemos investigar cuáles son las condiciones y correlatos neurales de tal posesión y de tal ejercicio. Ésa es la tarea de la neurociencia, que está descubriendo cada vez más cosas sobre todo ello. Pero sus descubrimientos no afectan en modo alguno a la verdad conceptual de que estas capacidades y su ejercicio en la percepción, el pensamiento y el sentimiento *son atributos de los seres humanos*, no de sus partes, en concreto, *no de su cerebro*. El ser humano es una unidad psicofísica, un animal que puede percibir, actuar intencionadamente, razonar y sentir emociones, un animal que utiliza el lenguaje, que no es meramente consciente, sino también autoconsciente; no un cerebro incrustado en el cráneo de un cuerpo. Sherrington, Eccles y Penfield concebían a los seres humanos como animales cuya mente, a la que imaginaban como portadora de los atributos psicológicos, está vinculada al cerebro. Suponer que el cerebro es el portador de los atributos psicológicos no supone avance alguno sobre esa concepción falsa.

Entre los neurocientíficos de hoy es muy habitual hablar de que el cerebro percibe, piensa, adivina o cree, o de que uno de sus hemisferios sabe cosas que el otro ignora. La justificación suele ser que ese tipo de discurso no es más que una mera forma de hablar. Pero esto es un grave error, porque la forma característica de explicación en la neurociencia cognitiva contemporánea consiste en adscribir atributos psicológicos al cerebro y sus partes *con el fin de explicar* la posesión de atributos psicológicos y el ejercicio (y las deficiencias en el ejercicio) de capacidades cognitivas por parte de los seres humanos.

La adscripción al cerebro de atributos psicológicos, en particular cognitivos y los cogitativos, es también fuente de graves confusiones adicionales, como bien demostramos. La neurociencia puede investigar las condiciones y los correlatos neurales de la adquisición, la posesión y el ejercicio de capacidades sensoriales por parte de los

animales. Puede descubrir cuáles son las requisitos neurales del ejercicio de las capacidades distintivamente humanas del pensamiento y razonamiento, de la memoria declarativa y la imaginación, de la emoción y la volición. Puede hacer todo ello mediante la paciente correlación inductiva entre los fenómenos neurales y la posesión y el ejercicio de las capacidades psicológicas, y entre lesiones neurales y deficiencias en las funciones mentales normales. Lo que *no puede* hacer es *sustituir* la amplia variedad de explicaciones psicológicas comunes de las actividades humanas en términos de razones, intenciones, propósitos, objetivos, valores, normas y convenciones, por unas explicaciones neurológicas.[6] Y *no puede* explicar cómo percibe o piensa un animal haciendo referencia al percibir o el pensar del cerebro, o de alguna de sus partes. Y es que no tiene sentido adscribir tales atributos psicológicos a algo que no sea el animal en su conjunto. Quien percibe es el animal, no partes de su cerebro, y son los seres humanos quienes piensan y razonan, no sus cerebros. El cerebro y sus actividades *hacen posible* que *nosotros*, no él, percibamos y pensemos, sintamos emociones y elaboremos proyectos que intentamos hacer realidad.

Si la respuesta inmediata de muchos neurocientíficos ante la acusación de confusión conceptual es afirmar que la adscripción de los predicados psicológicos al cerebro es una mera forma de hablar, su reacción al hecho demostrable de que sus teorías explicativas adscriben *de un modo nada trivial* capacidades psicológicas al cerebro es la de señalar que tal error es inevitable debido a las deficiencias del lenguaje. Discutimos esta falsa idea[7] [y] demostramos que los grandes descubrimientos de la neurociencia *no exigen* esta falsa forma de explicación, que lo que se ha descubierto puede describirse y explicarse perfectamente con el lenguaje de que disponemos. Y lo demostramos aduciendo los muy debatidos fenómenos que resul-

tan de una comisurotomía, descritos (en nuestra opinión, mal descritos) por Sperry, Gazzaniga y otros.[8]

En la Segunda Parte de *Philosophical Foundations of Neuroscience* (en adelante PFN) investigamos el uso de los conceptos de percepción, memoria, imaginería mental, emoción y volición en la actual teorización neurocientífica. Caso por caso demostramos que la falta de claridad conceptual, la incapacidad de prestar la atención adecuada a la estructura conceptual relevante, ha sido a menudo fuente de errores teóricos y base de inferencias equivocadas. Es un error, un error conceptual, suponer que la percepción es una cuestión de aprehender una *imagen* en la mente (Crick, Damasio, Edelman), formular una *hipótesis* (Helmholtz, Gregory) o generar *un modelo descriptivo en tres dimensiones* (Marr). Es confuso, una confusión conceptual, plantear el problema de la integración (*binding problem*) como el de combinar los datos de forma, color y movimiento para formar *la imagen* del objeto percibido (Crick, Kandel, Wurtz). Es inexacto, conceptualmente inexacto, suponer que la memoria siempre es memoria del pasado, o creer que los recuerdos pueden *almacenarse* en el cerebro en forma de la fuerza de las conexiones sinápticas (Kandel, Squire, Bennett). Y es una equivocación, una equivocación conceptual, suponer que investigar la sed, el hambre y el deseo sexual es una investigación sobre las emociones (Roles) o pensar que la función de las emociones es informarnos de nuestro estado visceral y músculo-esquelético (Damasio).

La reacción inmediata a tales observaciones críticas bien puede ser de indignación e incredulidad. ¿Cómo puede ser que una ciencia floreciente esté fundamentalmente equivocada? ¿Cómo podría haber una inevitable confusión conceptual en una ciencia consolidada? Si hay conceptos problemáticos, sin duda pueden sustituirse con facilidad por otros que no lo sean y que cumplan los mismos fines ex-

plicativos. Tales respuestas ponen de manifiesto una escasa comprensión de la relación entre la forma de la representación y los hechos representados, y una incomprensión de la naturaleza del error conceptual. También revelan una ignorancia de la historia de la ciencia en general y de la neurociencia en particular.

La ciencia no es más inmune al error conceptual y a la confusión que cualquier otra empresa intelectual. La historia de la ciencia está repleta de los despojos de teorías que no eran sólo factualmente erróneas, sino conceptualmente fracasadas. La teoría de la combustión de Stahl, por ejemplo, contenía el error *conceptual* de adscribir, en ciertas circunstancias, un peso negativo al flogisto —una idea que no tenía sentido en su marco conceptual, el de la física newtoniana—. Las famosas críticas de Einstein a la teoría del éter (el supuesto medio por el que se creía que se propagaba la luz) no iban dirigidas únicamente a los resultados de los experimentos de Michelson-Morley, que no habían conseguido detectar ningún efecto de movimiento absoluto, sino también a la confusión conceptual con respecto al movimiento relativo que suponía el papel adscrito al éter en la explicación de la inducción electromagnética. La neurociencia no ha sido una excepción, como mostramos en nuestro estudio histórico.[9] No cabe duda de que hoy es una ciencia *floreciente*, pero ello no la hace inmune a las confusiones y los enredos conceptuales. La cinemática newtoniana fue una ciencia floreciente, pero ello no evitó las confusiones conceptuales de Newton sobre la inteligibilidad de la acción a distancia, ni su desconcierto (no resuelto hasta Hertz) sobre la naturaleza de la fuerza. Del mismo modo, pues, el mayor logro de Sherrington, el de explicar la acción integradora de las sinapsis de la médula espinal y, consiguientemente, el de eliminar, de una vez por todas, la confusa idea de un «alma espinal», era perfectamente compatible con las confusiones conceptuales referentes

al «alma cerebral» o mente y su relación con el cerebro. Asimismo, los extraordinarios logros de Penfield en la identificación de localizaciones funcionales en la corteza, así como en desarrollar brillantes técnicas neuroquirúrgicas, eran perfectamente compatibles con las importantes confusiones sobre la relación entre la mente y el cerebro y sobre la «función superior del cerebro» (una idea que tomó prestada de Hughlings Jackson).

En pocas palabras, el desbarajuste conceptual *puede* coexistir con la ciencia floreciente. Tal vez parezca extraño. Si la ciencia puede florecer pese a tales confusiones conceptuales, ¿por qué iban a preocuparse por ellas los científicos? Los arrecifes ocultos no implican que los mares no sean navegables; sólo que son peligrosos. La cuestión discutible es cómo se manifiesta el avance sobre esos arrecifes. Las confusiones conceptuales se pueden manifestar de diferentes formas y en distintos puntos de la investigación. En algunos casos, es posible que la falta de claridad conceptual no afecte a la plausibilidad de las preguntas ni al provecho de los experimentos, sino únicamente a la interpretación de los resultados de éstos y sus implicaciones teóricas. Así, por ejemplo, Newton se embarco en la *Óptica* en busca de ideas sobre el carácter del color. Su investigación fue una contribución imperecedera a la ciencia. Pero su conclusión de que «los colores son sensaciones en el sensorium» demuestra que no consiguió alcanzar el tipo de comprensión que ansiaba. Porque los colores, sean lo que fueren, no son «sensaciones en el sensorium». Así pues, Newton, en la medida en que se preocupaba por la *interpretación* de los resultados de su investigación, tenía buenas razones para preocuparse por las confusiones conceptuales con las que trabajaba, ya que entorpecían el camino hacia una interpretación correcta.

Sin embargo, en otros casos la confusión conceptual también afecta a la investigación empírica. Es muy posible que las preguntas

equivocadas hagan trivial la investigación.[10] Y en el otro extremo, la errada construcción de los conceptos y las estructuras conceptuales conducirá en algunos casos a una investigación que en modo alguno será trivial, pero sí incapaz de demostrar aquello para lo que fue diseñada.[11] En estos casos, es posible que la ciencia no sea tan floreciente como aparenta ser, y se requiera una investigación conceptual para localizar los problemas y eliminarlos.

¿Son *inevitables* estas confusiones conceptuales? En absoluto. El único propósito de escribir este libro ha sido el de mostrar cómo evitarlas. Pero, naturalmente, no se pueden evitar si todo lo demás se deja intacto. *Se pueden* evitar, pero si se evitan dejarán de formularse determinados tipos de preguntas, ya que se pondrá de manifiesto que se asientan en un malentendido. Como dice Hertz en su magnífica introducción a sus *Principios de la mecánica*: «Cuando se eliminan esas lamentables contradicciones, [...] nuestra mente, libre ya del desconcierto, dejará de formular preguntas espurias». Del mismo modo, dejarán de hacerse ciertos tipos de inferencias a partir de un determinado cuerpo de estudios empíricos, ya que se revelará que tienen poco o nada que ver con la materia que se pretendía esclarecer, aunque puedan tener algo que ver con *otra cosa*.

Si existen conceptos problemáticos, ¿se pueden reemplazar por otros que cumplan la misma función explicativa? El científico siempre es libre de introducir nuevos conceptos si considera que los existentes son inadecuados o no están lo bastante elaborados. Pero nuestra preocupación en este libro no radica en el uso de nuevos conceptos técnicos, sino en el uso indebido de viejos conceptos no técnicos, los conceptos de mente y cuerpo, pensamiento e imaginación, sensación y percepción, conocimiento y memoria, movimiento voluntario y conciencia y autoconciencia. Nada hay de inadecuado en estos conceptos en lo que se refiere a los propósitos a los que sirven. No

hay razón para pensar que se deban sustituir en los contextos que nos ocupan. El problema está en la interpretación errónea que de ellos hacen los neurocientíficos y los malentendidos que se siguen de ella. Tales malentendidos pueden remediarse mediante una explicación correcta de la categoría lógico-gramatical de los conceptos en cuestión. Y esto es justamente lo que hemos intentado proporcionar.

Si admitimos que los neurocientíficos tal vez no estén utilizando estos conceptos comunes como los utiliza el hombre de la calle, ¿con qué derecho puede pretender corregirlos la filosofía? ¿Cómo puede la filosofía cuestionar con tanta seguridad la claridad y la coherencia de los conceptos empleados por científicos competentes? ¿Cómo puede estar la filosofía en situación de proclamar que ciertas afirmaciones de sofisticados neurocientíficos *no tienen sentido*? En las páginas que siguen resolveremos estas dudas metodológicas. Pero es posible que cierta clarificación inicial en este punto disipe algunas de ellas. La verdad y la falsedad son a la ciencia lo que el sentido y el sinsentido son a la filosofía. El error de observación y teórico deriva en falsedad; el error conceptual, en falta de sentido. ¿Cómo se pueden investigar los límites del sentido? Sólo mediante el examen del uso de las palabras. El sinsentido a menudo aparece cuando una expresión se emplea en contra de las reglas que rigen su uso. La expresión en cuestión puede ser una expresión corriente y no técnica, en cuyo caso las reglas para su uso pueden deducirse de su empleo habitual y de las explicaciones estándar de su significado. O puede ser un término técnico de una disciplina, en cuyo caso las reglas para su uso deben deducirse de la introducción que el teórico hace del término y de las explicaciones que ofrece del uso estipulado. Ambos tipos de términos se pueden usar indebidamente, y cuando eso ocurre se produce el sinsentido, una combinación de palabras que está excluida del lenguaje. Y es que o bien no se ha estipulado nada sobre

lo que el término significa en el aberrante contexto en cuestión, o bien esta combinación de palabras está de hecho excluida por una regla que especifica que no existe nada así (por ejemplo, que no existe nada que sea el «este del Polo Norte»), que ésta es una combinación de palabras que no tiene uso. El sinsentido también se suele generar cuando a una expresión se le da un uso nuevo, quizá técnico o casi técnico, y el nuevo uso se solapa inadvertidamente con el antiguo, por ejemplo cuando a partir de proposiciones que contienen el nuevo término se hacen inferencias qué sólo son legítimas en relación con el antiguo. Identificar tales transgresiones de los límites del sentido es tarea del crítico conceptual. No basta, por supuesto, con mostrar que un determinado científico ha empleado un término en contra de su uso habitual, pues es muy posible que lo utilice en un nuevo sentido. El crítico debe demostrar que el científico pretende usar el término en su sentido acostumbrado, pero no lo hace; o que pretende usarlo en un sentido nuevo, pero sin percatarse de que ha solapado el nuevo sentido y el antiguo. Siempre que sea posible, el descarriado científico será sentenciado por sus propias palabras.[12]

El último error del que queremos advertir es la idea de que nuestras reflexiones son irremisiblemente negativas. Podría creerse que nuestra única tarea es desaprobar. Nuestra obra puede aparecer bajo una perspectiva superficial como una empresa meramente destructiva, que no proporciona asistencia ni nuevos caminos por los que avanzar. Peor aún, podría parecer que se está intentando enfrentar a la filosofía y la neurociencia cognitiva. *Nada más lejos de la verdad.*

Hemos escrito este libro desde la admiración ante los logros de la neurociencia del siglo XX, y el deseo de colaborar con la disciplina. Pero las *únicas* formas en que una investigación conceptual puede ayudar a una disciplina empírica son la identificación del error conceptual (si lo hay) y la provisión de un mapa que ayude a evitar

que los investigadores empíricos se desvíen del buen camino del sentido. Así pues, cada una de nuestras investigaciones tiene dos aspectos. Por un lado, hemos intentado identificar los problemas y los enredos conceptuales presentes en importantes teorías actuales de la percepción, la memoria, la imaginación, la emoción y la volición. Además, sostenemos que gran parte de los escritos sobre la naturaleza de la conciencia y la autoconciencia está plagada de dificultades conceptuales. Este aspecto de nuestras investigaciones es, sin duda, negativo y crítico. Por otro lado, nos hemos esforzado, caso por caso, por ofrecer una representación clara del campo conceptual de cada uno de los conceptos problemáticos. Éste es un empeño constructivo. Confiamos en que estas visiones de conjunto conceptuales ayuden a los neurocientíficos en sus reflexiones previas al diseño de experimentos. Sin embargo, no puede ser tarea de una investigación conceptual formular hipótesis empíricas que puedan resolver los problemas empíricos a los que se enfrentan los científicos. Lamentarse de que una investigación filosófica sobre la neurociencia cognitiva no haya aportado una nueva teoría neurocientífica es como quejarse ante el matemático que ha demostrado un teorema nuevo de que no sea una nueva teoría física.

PHILOSOPHICAL FOUNDATIONS OF NEUROSCIENCE

Fragmento del capítulo 3

MAXWELL BENNETT y PETER HACKER

3.1 CONFUSIONES MEREOLÓGICAS EN NEUROCIENCIA COGNITIVA

Adscripción de atributos psicológicos al cerebro

Las figuras más destacadas de las dos primeras generaciones de neurocientíficos modernos del cerebro fueron fundamentalmente cartesianos. Al igual que Descartes, distinguían la mente del cerebro y adscribían atributos psicológicos a la mente. En consecuencia, la atribución de tales predicados a los seres humanos era derivativa, como en la metafísica cartesiana. Sin embargo, la tercera generación de neurocientíficos rechazó el dualismo de sus maestros. En el proceso de explicar la posesión de atributos psicológicos por parte de los seres humanos, adscribían tales atributos no a la mente, sino al cerebro o a sus partes.

Los neurocientíficos presumen que el cerebro tiene una amplia variedad de capacidades cognitivas, cogitativas, perceptivas y volitivas. Francis Crick afirma:

> Lo que ves no es *realmente* lo que hay; es lo que tu cerebro *cree* que hay [...] El cerebro hace la mejor interpretación que puede de acuerdo

con su experiencia previa y la información limitada y ambigua que los ojos le proporcionan [...] El cerebro combina la información que las muy diferentes características de la escena visual ofrecen (aspectos de la forma, el color, el movimiento, etc.) y se asienta en la interpretación más verosímil de todos esos indicios tomados en su conjunto [...] Lo que el cerebro debe construir es una interpretación de múltiples niveles de la escena visual [...] [Este proceso de reconstrucción] permite al cerebro conjeturar una imagen completa a partir de sólo información parcial —una habilidad muy útil.[1]

Así pues, el cerebro *tiene experiencias*, *cree* cosas, *interpreta* pistas *a partir de la información* que se le facilita y *hace conjeturas*. Gerald Edelman sostiene que las estructuras internas del cerebro «categorizan, disciernen y recombinan las diversas actividades del cerebro que tienen lugar en distintos tipos de mapeos globales», y que el cerebro «relaciona de forma recursiva las secuencias semánticas con las fonológicas, y luego genera unas correspondencias sintácticas, no a partir de reglas preexistentes, sino tratando las reglas que se generan en la memoria como objetos de manipulación conceptual».[2] Por lo tanto, el cerebro *categoriza*, en efecto, «categoriza sus propias actividades (en particular sus categorizaciones perceptivas)» y *manipula conceptualmente las reglas*. Colin Blakemore sostiene:

Parece que nos vemos llevados a decir que tales neuronas [tal como reaccionan de un modo altamente específico a, por ejemplo, la orientación lineal] poseen conocimientos. Tienen inteligencia, pues son capaces de calcular la probabilidad de acontecimientos externos, unos acontecimientos que son importantes para el animal en cuestión. Y el cerebro adquiere sus conocimientos mediante un proceso análogo al razonamiento inductivo del método científico clásico. Las neuronas presentan al cerebro unos argumentos basados en las características con-

cretas que detectan, unos argumentos sobre los que el cerebro construye sus hipótesis de percepción.[3]

Así pues, el cerebro *sabe* cosas, *razona* de forma inductiva, *construye hipótesis* basadas en argumentos, y las neuronas que lo constituyen son *inteligentes*, saben *calcular probabilidades* y *ofrecen argumentos*. J. Z. Young comparte en gran medida esta concepción. Afirma que «podemos considerar todo acto de ver como una búsqueda continua de las respuestas a preguntas formuladas por el cerebro. Las señales que proceden de la retina constituyen "mensajes" que transmiten esas respuestas. A continuación, el cerebro utiliza esta información para construir una hipótesis adecuada sobre lo que hay».[4] Por consiguiente, el cerebro *formula preguntas, busca respuestas* y *construye hipótesis*. Antonio Damasio afirma que «nuestro cerebro suele decidir bien, en unos segundos o unos minutos, en función del marco temporal que establecemos como apropiado para la meta que deseamos alcanzar, y si puede hacer tal cosa, tiene que realizar tan maravilloso trabajo con algo más que la razón pura»,[5] y Benjamin Libet señala que «el cerebro "decide" iniciar o, al menos, preparar el inicio del acto antes de que haya una conciencia subjetiva que merezca reseñarse de que ha tenido lugar tal decisión».[6] Es decir, el cerebro *decide*, o al menos «decide», e *inicia la acción voluntaria*.

Los psicólogos coinciden. J. P. Frisby sostiene que «en el cerebro debe haber una descripción simbólica del mundo exterior, una descripción proyectada en símbolos que representan los diversos aspectos del mundo de los que la vista nos hace conscientes».[7] De modo que hay *símbolos en el cerebro*, y éste *usa símbolos* y, presumiblemente, los *comprende*. Richard Gregory entiende la visión como «probablemente la más compleja de todas las actividades del cerebro: recurre a sus reservas de datos de la memoria; exige sutiles cla-

sificaciones, comparaciones y decisiones lógicas para que los datos sensoriales se conviertan en percepción».[8] Por lo tanto, el cerebro *ve, hace clasificaciones, comparaciones* y toma *decisiones.* Y los científicos cognitivos opinan lo mismo. David Marr sostiene que «nuestro cerebro tiene que ser de algún modo capaz de *representar* [...] *la información* [...] Por consiguiente, el estudio de la visión debe incluir [...] también una investigación sobre la naturaleza de las *representaciones internas* por las que *captamos esta información* y la *hacemos accesible como base para decisiones* acerca de nuestros pensamientos y nuestras acciones».[9] Y Philip Johnson-Laird sugiere que el cerebro «tiene acceso a un modelo parcial de sus propias capacidades» y posee la «maquinaria recursiva para insertar modelos dentro de modelos»; la conciencia, afirma, «es la propiedad de una clase de algoritmos paralelos».[10]

Cuestionar la inteligibilidad de la adscripción de atributos psicológicos al cerebro

Con este amplio consenso sobre la forma correcta de pensar acerca de las funciones del cerebro y acerca de la explicación de los requisitos causales para que los seres humanos posean y ejerzan sus capacidades naturales de pensamiento y percepción, uno se siente inclinado a dejarse llevar por entusiastas proclamas —de nuevos campos del conocimiento conquistados, nuevos misterios desvelados—.[11] Pero debemos tomarnos las cosas con calma, y detenernos a pensar. Sabemos lo que sienten los seres humanos al experimentar cosas, ver cosas, saber o creer cosas, tomar decisiones, interpretar datos equívocos, conjeturar y formular hipótesis. Comprendemos lo que para los seres humanos es razonar por inducción, calcular las pro-

babilidades, exponer argumentos, clasificar y categorizar las cosas con que se topan en su experiencia. Planteamos preguntas y buscamos respuestas, utilizando un simbolismo, nuestro lenguaje, en cuyos términos nos representamos las cosas. ¿Pero sabemos lo que es para *el cerebro* ver u oír, tener experiencias, saber o creer algo? ¿Tenemos alguna idea de lo que sería para *el cerebro* tomar una decisión? ¿Intuimos qué es para el cerebro (y no digamos para una neurona) *razonar* (de modo inductivo o deductivo), *calcular las probabilidades*, *exponer argumentos*, *interpretar datos* y *formular hipótesis* a partir de sus interpretaciones? Podemos observar si una persona ve una cosa u otra; nos fijamos en su conducta y le hacemos preguntas. ¿Pero qué sería observar si un cerebro ve algo, en oposición a observar el cerebro de *una persona* que ve algo? Reconocemos cuándo una persona hace una pregunta y cuándo otra la responde. ¿Pero tenemos idea de lo que sería para un cerebro hacer una pregunta o responderla? Todos éstos son atributos de los seres humanos. ¿Es un *descubrimiento* nuevo que el cerebro también realiza este tipo de actividades humanas? ¿O se trata de una innovación lingüística, introducida por neurocientíficos, psicólogos y científicos cognitivos, que extienden el uso corriente de estas expresiones psicológicas por buenas razones teóricas? O lo que es más inquietante, ¿se trata de una confusión conceptual? ¿Podría ser que simplemente no es el caso que el cerebro piense o sepa, vea u oiga, crea o adivine, posea y utilice información, formule hipótesis, es decir, que estas formas de combinar palabras no tengan sentido? Pero si no es el caso, ¿por qué tantos distinguidos científicos pensaron que esas frases, así empleadas, sí tienen sentido?

La cuestión de si se pueden adscribir de forma inteligible atributos psicológicos al cerebro es una cuestión filosófica y por lo tanto conceptual, no científica

La pregunta a la que nos enfrentamos es una pregunta filosófica, no científica. Requiere un esclarecimiento conceptual, no una investigación experimental. No se puede investigar experimentalmente si los cerebros piensan o no, creen o no, adivinan o no, razonan o no, formulan hipótesis o no, etc. mientras no sepamos qué significaría que el cerebro hiciera tales cosas, es decir, mientras no tengamos claro el significado de esas frases y sepamos cuándo se *considera* (si es que algo se considera) que un cerebro hace todo eso, y qué tipo de pruebas avalan la adscripción de tales atributos al cerebro. (No se pueden buscar los polos de la Tierra mientras no se sepa qué es un polo, es decir, qué significa el término «polo», y también cuándo se considera que se ha descubierto un polo de la Tierra. De no ser así, uno podría embarcarse, como Winnie-the-Pooh, en una expedición hacia el Polo Este.) La pregunta a debate es: ¿tiene sentido adscribir tales atributos al cerebro? ¿Existe eso del pensar, el creer, etc. del cerebro? (¿Existe eso del Polo Este?)

En las *Investigaciones filosóficas*, Wittgenstein hace una profunda observación que guarda una relación directa con lo que nos ocupa: «*Sólo del ser humano y de lo que se parece a un ser humano (se comporta como tal) se puede decir: tiene sensaciones; ve, es ciego; oye, es sordo; es consciente o inconsciente*».[12] Aquí está el epítome de las conclusiones a las que llegaremos en nuestra investigación. Dado el acostumbrado laconismo del autor, es necesario desarrollar la idea y dilucidar sus ramificaciones.

No se trata de una cuestión factual. No es un hecho que únicamente se pueda decir de los seres humanos y de lo que se compor-

ta como un ser humano que son el sujeto de estos predicados psicológicos. Si así fuera, podría sin duda ser un descubrimiento, obra reciente de los neurocientíficos, que los cerebros también ven y oyen, piensan y creen, formulan y responden preguntas. Tal descubrimiento demostraría, sin duda, que tales cosas se pueden decir no sólo del ser humano y de lo que se comporta como tal. Sería algo asombroso, y desearíamos saber más al respecto. Querríamos saber cuáles serían las pruebas de ese notable descubrimiento. Pero, evidentemente, no es éste el caso. La atribución de atributos psicológicos al cerebro no está avalada por ningún descubrimiento neurocientífico que demuestre que, contrariamente a nuestras anteriores convicciones, los cerebros realmente piensan y razonan, tal como nosotros mismos hacemos. Los neurocientíficos, psicólogos y científicos cognitivos que adoptan estas formas de descripción no lo hacen como resultado de unas *observaciones* que demuestren que el cerebro piensa y razona. Susan Savage-Rambaugh ha reunido pruebas concluyentes que demuestran que los chimpancés bonobo, debidamente entrenados y enseñados, pueden plantear y responder preguntas, razonar de forma rudimentaria, dar y obedecer órdenes, etc. La prueba está en su comportamiento, en lo que hacen en sus interacciones con nosotros (incluida la forma que tienen de emplear símbolos). Este descubrimiento es realmente asombroso, porque nadie imaginaba que los simios pudieran adquirir tales capacidades. Pero sería absurdo pensar que la atribución al cerebro de atributos cognitivos y cogitativos se asienta en pruebas comparables. Sería absurdo porque ni siquiera sabemos qué es lo que demostraría que el cerebro posee tales atributos.

La equivocada adscripción de los atributos psicológicos al cerebro es una forma degenerada de cartesianismo

Entonces, ¿por qué esta forma de descripción y las formas de explicación que la acompañan y que dependen de ella se adoptaron *sin argumentos ni reflexión*? Sospechamos que la respuesta es que fue consecuencia de una adhesión no meditada a una forma mutante de cartesianismo. Atribuir predicados psicológicos a la mente, y sólo por derivación al ser humano, fue una característica del dualismo cartesiano. En sus reflexiones sobre la relación entre sus descubrimientos neurológicos y las capacidades perceptivas y cognitivas humanas, Sherrington y sus alumnos Eccles y Penfield permanecieron fieles a una forma de dualismo. Sus sucesores rechazaron el dualismo, con mucha razón. Pero los neurocientíficos del cerebro de la tercera generación aplicaron irreflexivamente al cerebro los mismos predicados que los dualistas atribuían a la mente inmaterial. No era más que un corolario aparentemente inocuo del rechazo de los neurocientíficos del dualismo sustancial cartesiano. Estos científicos procedieron a explicar las capacidades perceptivas y cognitivas humanas y su ejercicio a partir del ejercicio, por parte del cerebro, de *sus propias* capacidades cognitivas y perceptivas.

La adscripción de atributos psicológicos al cerebro carece de sentido

Nuestra tesis es que esta aplicación de los predicados psicológicos al cerebro *no tiene sentido*. No es que de hecho el cerebro no piense, no formule hipótesis ni decida, no vea ni oiga, no formule ni res-

ponda preguntas, sino más bien que no tiene sentido atribuir tales predicados *o sus negaciones* al cerebro. El cerebro ni ve *ni es ciego*, del mismo modo que los palos y las piedras no están despiertos, *pero tampoco están dormidos*. El cerebro no oye, pero no es sordo, no más de lo que puedan serlo los árboles. El cerebro no toma decisiones, pero tampoco es indeciso. Sólo lo que *puede* decidir puede ser indeciso. Así, el cerebro tampoco puede ser consciente, sólo la criatura de la que es cerebro puede ser consciente, o inconsciente. *El cerebro no es un sujeto lógicamente apropiado de predicados psicológicos*. Sólo del ser humano y de lo que *se comporta como tal* se puede decir de forma inteligible y literal que ve o es ciego, oye o es sordo, formula preguntas o se abstiene de preguntar.

Así pues, el asunto aquí es conceptual. No tiene sentido atribuir predicados psicológicos (o sus negaciones) al cerebro, excepto de forma metafórica o metonímica. La combinación de palabras resultante no dice algo que sea falso, más bien no dice nada en absoluto, pues carece de sentido. Los predicados psicológicos son predicados que se aplican esencialmente al animal viviente en su conjunto, no a sus partes. No es el ojo (y mucho menos el cerebro) quien ve, sino que *nosotros* vemos *con* nuestros ojos (y no vemos *con* nuestros cerebros, aunque, sin un cerebro cuyo sistema visual funciona con normalidad, no veríamos). Del mismo modo, quien oye no es el oído, sino el animal de quien es el oído. Los órganos de un animal son partes del animal, y los predicados psicológicos son atribuibles a todo el animal, no a sus partes constituyentes.

La atribución por parte de los neurocientíficos de los atributos psicológicos al cerebro se puede denominar «la falacia mereológica» de la neurociencia

La mereología es la lógica de las relaciones entre la parte y el todo. Llamaremos «*falacia mereológica*» en neurociencia al error de los neurocientíficos de atribuir a las *partes* constituyentes de un animal atributos lógicamente aplicables sólo al animal como un *todo.*[13] Llamaremos «*principio mereológico*» en neurociencia al principio de que los predicados psicológicos aplicables únicamente a un ser humano (u otro animal) en su totalidad no se pueden aplicar de modo inteligible a sus partes, por ejemplo al cerebro.[14] Se puede decir que los seres humanos, pero no su cerebro, son reflexivos o irreflexivos; se puede decir de los animales, pero no de su cerebro, y mucho menos de los hemisferios de éste, que ven, oyen, huelen y saborean cosas; podemos decir de las personas, pero no de su cerebro, que toman decisiones o que son indecisorias.

Conviene señalar que existen muchos predicados que se *pueden* aplicar tanto a un determinado todo (en particular, un ser humano) como a sus partes, y cuya aplicación en un caso se puede deducir de su aplicación en el otro. Un hombre puede estar quemado por el sol y su cara puede estar quemada por el sol; puede estar todo él temblando, de manera que también sus manos estarán temblando. Del mismo modo, a veces la aplicación de un predicado a un ser humano se extiende a las partes del cuerpo humano, por ejemplo cuando decimos que un hombre agarró el picaporte, o también que su mano agarró el picaporte; que resbaló o que su pie resbaló. No hay aquí nada lógicamente inadecuado. Pero los predicados psicológicos se aplican paradigmáticamente al *ser humano (o animal) como un todo*, y *no* al cuerpo y sus partes. Hay unas pocas excepciones, por ejemplo

la aplicación de los verbos que denotan sensación, como «doler» o «hacer daño», a partes del cuerpo, como en los casos de «Me duele la mano», «Me haces daño en la mano».[15] Pero la variedad de predicados psicológicos que nos interesan, es decir, aquellos que los neurocientíficos, los psicólogos y los científicos cognitivos invocan en su empeño por explicar las capacidades humanas y su ejercicio, no tienen una aplicación literal a las partes del cuerpo. En particular, no tienen una aplicación inteligible al cerebro.

3.2 Dudas Metodológicas

Objeciones metodológicas a la acusación de que los neurocientíficos incurren en una falacia mereológica

Si una persona atribuye un predicado a una entidad a la que el predicado en cuestión no es aplicable lógicamente, y se le hace tal observación, sólo cabe esperar que insista con indignación en que «no era eso lo que quería decir». Al fin y al cabo, podría argüir esa persona, dado que un sinsentido es una combinación de palabras que no dice nada y que no describe ningún estado de cosas posible, él obviamente no *quería decir* un sinsentido —uno *no puede* querer decir un sinsentido, ya que no hay nada que querer decir—. Así que no hay que entender sus palabras en su significado habitual. Tal vez las expresiones problemáticas se usaron en un sentido especial, y en realidad son meramente homónimas; o eran *extensiones* analógicas del uso acostumbrado, tan comunes en ciencia; o se usaron en un sentido metafórico o figurativo. Si se puede recurrir a tales vías de escape, la acusación de que los neurocientíficos han sido víctimas de la falacia mereológica está injustificada. Aunque hacen uso del mismo voca-

bulario psicológico que emplea el hombre de la calle, lo están empleando de forma distinta. De modo que las objeciones al uso de los neurocientíficos basadas en el uso corriente de estas expresiones son irrelevantes.

Sin embargo, las cosas no son tan simples. Es evidente que la persona que adscribe erróneamente un predicado de la manera en cuestión no pretende proferir una combinación de palabras que carece de sentido. Pero el hecho de que no quiera *proferir* un sinsentido no garantiza que no lo haga. Aunque naturalmente va a insistir en que «no quería decir eso», que el predicado en cuestión no se usaba en su sentido habitual, la insistencia de esa persona no es la que tiene la última palabra. La autoridad última sobre el asunto es *el propio razonamiento del hablante.* Debemos fijarnos en las consecuencias que saca de sus propias palabras, y serán las inferencias que haga las que demostrarán si estaba usando el predicado en un sentido nuevo o si hacía de él un mal uso. Si hay que condenarle, tiene que ser por su propia boca.

Así pues, observemos las propuestas vías de escape con las que se pretende demostrar que los neurocientíficos y los científicos cognitivos no son culpables de los errores de los que les hemos acusado.

Primera objeción (Ullmann): los predicados psicológicos así usados son homónimos de predicados psicológicos corrientes, y tienen un significado técnico diferente

En primer lugar, se podría señalar que, en efecto, los neurocientíficos emplean homónimos, cuyo significado es completamente distinto. Nada tiene de inusual, y mucho menos de problemático, que los científicos introduzcan una nueva forma de hablar bajo la presión

de una teoría nueva. Si esto confunde a los lectores ignorantes, la confusión se puede resolver con facilidad. Evidentemente, los cerebros no piensan, creen, infieren, interpretan ni formulan hipótesis de forma literal, sino que piensan*, creen*, infieren*, interpretan* o formulan hipótesis*. No tienen ni construyen representaciones simbólicas, sino representaciones* simbólicas.[16]

Segunda objeción (Gregory): los predicados psicológicos así usados son extensiones analógicas de las expresiones corrientes

En segundo lugar, se podría señalar que los neurocientíficos extienden por analogía el uso corriente del vocabulario relevante, como a menudo se ha hecho en la historia de la ciencia, por ejemplo en la extensión analógica de la hidrodinámica en la teoría de la electricidad. Por tanto, poner objeciones a la atribución de predicados psicológicos al cerebro basándose en que en el habla corriente tales predicados sólo son aplicables al animal como un todo significaría mostrar una forma de *inercia semántica.*[17]

Tercera objeción (Blakemore): la adscripción que los neurocientíficos hacen de los atributos psicológicos al cerebro es figurativa o metafórica, pues saben perfectamente bien que el cerebro no piensa ni usa mapas

Finalmente, se podría argumentar que los neurocientíficos *realmente* no piensan que el cerebro razone, discuta, formule y responda preguntas del mismo modo que nosotros lo hacemos. No piensan de

verdad que el cerebro interprete pistas, haga conjeturas o contenga símbolos que describen el mundo exterior. Y aunque hablan de la existencia de «mapas» en el cerebro y de que éste contiene «representaciones internas», no emplean estas palabras en su sentido habitual o vulgar. Es una manera de hablar figurativa y metafórica, a veces incluso una licencia poética.[18] Por consiguiente, a los neurocientíficos no los confunden en absoluto esas formas de hablar; saben perfectamente bien lo que quieren decir, pero carecen de las palabras para decirlo, a no ser que lo hagan de modo metafórico o figurativo.

Réplica a la objeción de que los neurocientíficos utilizan el vocabulario psicológico en un sentido técnico especial

Por lo que se refiere al mal uso del vocabulario psicológico implícito en la atribución de predicados psicológicos al cerebro, todas las pruebas apuntan al hecho de que los neurocientíficos no emplean estos términos en un sentido especial. Las expresiones psicológicas que emplean, lejos de ser nuevos usos homónimos, se invocan en su sentido acostumbrado, de lo contrario los neurocientíficos no inferirían de ellas lo que infieren. Cuando Crick afirma que «lo que vemos no es *realmente* lo que hay; es lo que nuestro cerebro *cree* que hay», es importante que asuma que «cree» tiene sus connotaciones normales —que no significa lo mismo que cierto término nuevo «cree*»—. Porque forma parte de la teoría de Crick que la creencia es el resultado de una *interpretación* basada en la *experiencia* y la *información* previas (y no el resultado de una interpretación* basada en la experiencia* y la información* previas). Cuando Semir Zeki subraya que la adquisición del conocimiento es una «función primordial del cerebro»,[19] se refiere al conocimiento (no al conocimiento*), de lo contrario no

pensaría que es tarea de la neurociencia del futuro resolver los problemas de la epistemología (sino sólo, presumiblemente, de la epistemología*). Asimismo, cuando Young afirma que el cerebro contiene conocimiento e información, que están codificados en el cerebro «del mismo modo que el conocimiento está registrado en libros u ordenadores»,[20] se refiere al conocimiento, no al conocimiento* —ya que son el conocimiento y la información, no el conocimiento* y la información*, lo que se puede registrar en libros y ordenadores—. Cuando Milner, Squire y Kandel hablan de «memoria declarativa», aclaran que esta expresión significa «lo que normalmente se quiere decir con el término "memoria"»,[21] pero luego afirman que los recuerdos asociados a tal memoria (no memoria*) se «almacenan en el cerebro». Esto presupone que tiene sentido hablar de la memoria (en el sentido corriente de la palabra) como la actividad de almacenar recuerdos *en el cerebro.*[22]

Réplica a Ullmann: David Marr sobre las «representaciones»

La acusación de incurrir en la falacia mereológica no se puede refutar tan fácilmente. Pero puede parecer que Simon Ullman se asienta en bases más sólidas cuando habla de representaciones interiores y representaciones simbólicas (así como de mapas) *en el cerebro.* Si «representación» no significa lo que normalmente significa, si «simbólico» no tiene nada que ver con los símbolos, entonces, en efecto, puede ser inocuo decir que en el cerebro hay representaciones interiores y simbólicas. (Y si los «mapas» no tienen nada que ver con los atlas, sino sólo con operaciones de *mapeo*, entonces también puede ser inocuo decir que en el cerebro hay mapas.) Multiplicar los homónimos es extremadamente desaconsejable, pero no implica ne-

cesariamente incoherencia conceptual, *siempre y cuando los científicos que emplean estos términos de esa forma no olviden que los términos no tienen su significado habitual.* Lamentablemente, lo habitual es que lo olviden y *procedan a cruzar el uso nuevo con el antiguo*, cayendo así en una incoherencia. Ullman, en defensa de Marr, insiste (de forma perfectamente correcta) en que ciertos acontecimientos del cerebro se pueden considerar representaciones* de la profundidad, la orientación o la reflectancia,[23] es decir que ciertos patrones de activación neuronal se pueden correlacionar con características del campo visual (y hablar de los primeros como «representaciones*» de las segundas). Pero es evidente que esto no es todo lo que Marr quiere decir. Sostiene que los sistemas numerales (los numerales romanos o árabes, la notación binaria) son representaciones. Sin embargo, tales notaciones no tienen nada que ver con correlaciones causales, sino con convenciones representacionales. Sostiene que «una representación de la forma sería un esquema formal descriptivo de algunos aspectos de la forma, junto con reglas que especifican cómo aplicar el esquema a una forma particular»;[24] que un esquema formal es «un conjunto de símbolos y las reglas para combinarlos»;[25] y que «por lo tanto, una representación no es en absoluto una idea extraña; todos utilizamos representaciones continuamente. Sin embargo, la noción de que uno puede captar cierto aspecto de la realidad haciendo una descripción de él y usando para ello un símbolo, y que hacer tal cosa puede ser útil, me parece una idea poderosa y fascinante».[26] Pero el sentido en que «utilizamos las representaciones continuamente», *en que las representaciones son símbolos que se rigen por reglas, y en que se usan para describir cosas*, es el sentido *semántico* de «representación», no un sentido homonímico nuevo basado en una correlación causal . Marr cae en su propia trampa.[27] En realidad, confunde las representaciones* de Ullman, que *son* correlatos

causales, con las representaciones, que son símbolos o sistemas de símbolos con una sintaxis y un significado determinados por convención.

Réplica a Ullman: Young sobre los «mapas» y Frisby sobre las «representaciones simbólicas»

De modo similar, sería un error, por otra parte inocuo, hablar de mapas del cerebro cuando lo que se quiere decir es que ciertas características del campo visual se pueden mapear sobre las descargas de grupos de células de la corteza visual primaria o estriada. Pero de ahí no se puede deducir, como hace Young, que el cerebro hace uso de sus mapas cuando formula hipótesis sobre lo que es visible. También sería inocuo decir que existen representaciones simbólicas en el cerebro, siempre y cuando «simbólico» no tenga nada que ver con el sentido semántico, sino que signifique únicamente «significado natural» (como en «el humo significa fuego»). Pero entonces no podemos añadir, como hace Frisby, que «en el cerebro debe de haber una descripción simbólica del mundo exterior, una descripción formulada en símbolos que se refieren a los varios aspectos del mundo de los que la vista nos hace conscientes».[28] Y es que este uso de «símbolo» es evidentemente semántico. Porque si el humo significa fuego, en el sentido de que es un signo *de* fuego (una indicación correlacionada con el fuego por inducción), no es un signo *del* fuego. El humo que se eleva en una distante colina no es una descripción del fuego formulada en símbolos, ni la activación de las neuronas de la corteza estriada «visual» es una descripción simbólica de los objetos del campo visual, aunque el neurocientífico pueda hacer inferencias empíricas sobre qué es visible para un animal a partir del conocimiento que

posee sobre qué células están activadas en su corteza estriada «visual». La activación de células en V1 puede ser el signo de una figura con determinadas orientaciones lineales del campo visual del animal, pero no *significa* nada, no es un símbolo, y no describe nada.

Réplica a la segunda objeción de que, al adscribir atributos psicológicos al cerebro, los neurocientíficos no cometen la falacia mereológica, sino que simplemente extienden analógicamente el vocabulario psicológico

Otra manera de eludir la acusación de que las descripciones que los neurocientíficos hacen de sus descubrimientos suelen transgredir los límites del sentido es la idea de que el uso neurocientífico, lejos de ser conceptualmente incoherente, es innovador, pues extiende el vocabulario psicológico en maneras novedosas. Es cierto, sin duda, que las analogías son una fuente de comprensión científica. La analogía hidrodinámica demostró ser provechosa en el desarrollo de la teoría de la electricidad, pese a que la corriente eléctrica no fluye del mismo modo que el agua, y que el cable eléctrico no es una tubería de ningún tipo. Lo que se puede discutir es si la aplicación del vocabulario psicológico al cerebro se debe entender como analógica.

Las perspectivas de que sea así no parecen buenas. La aplicación de expresiones psicológicas al cerebro no forma parte de una compleja teoría repleta de relaciones funcionales y matemáticas que se pueden expresar mediante leyes cuantificables como las que debe haber en la teoría de la electricidad. Parece que se necesita algo mucho más flexible. Así pues, es cierto que los psicólogos, siguiendo a Freud y a otros, han extendido los conceptos de creencia, deseo y motivo para hablar de creencias, deseos y motivos *inconscientes*. Cuando estos

conceptos son sometidos a tal extensión, aparece algo nuevo que necesita una explicación. Las expresiones recién extendidas ya no admiten las mismas posibilidades combinatorias que antes. Tienen un significado diferente, que guarda una importante relación con el anterior, y que requiere una explicación. La relación entre una creencia (consciente) y una creencia inconsciente, por ejemplo, no es igual que la relación entre una silla visible y una silla oculta, no es «simplemente como una creencia consciente sólo que inconsciente», sino más bien como la relación entre $\sqrt{1}$ y $\sqrt{-1}$. Pero cuando neurocientíficos como Sperry y Gazzaniga dicen que el hemisferio izquierdo toma decisiones, genera interpretaciones, sabe, observa y explica cosas, resulta evidente por lo que añaden a continuación que a estas expresiones psicológicas no se les ha dado un significado nuevo. De otro modo, no dirían que un hemisferio del cerebro es «un sistema consciente por derecho propio, que percibe, piensa, recuerda, razona, desea y siente, *todo a un nivel característicamente humano*».[29]

No es la inercia semántica lo que motiva nuestra afirmación de que los neurocientíficos incurren en diversas formas de incoherencia. Es más bien el reconocimiento de las exigencias de la lógica de las expresiones psicológicas. Los predicados psicológicos sólo se pueden predicar de un animal entero, no de sus partes. No se han establecido convenciones para determinar lo que se puede significar con la atribución de tales predicados a una parte de un animal, en concreto a su cerebro. De modo que la aplicación de tales predicados al cerebro o a sus hemisferios sobrepasa los límites del sentido. Las afirmaciones resultantes no son falsas, ya que para decir que algo es falso debemos tener cierta idea de lo que supondría que fuera verdadero —en este caso, deberíamos saber qué supondría que el cerebro pensara, razonara, viera y oyera, etc. y haber descubierto que, de hecho, el cerebro no hace tales cosas—. Pero no tenemos esa idea, y estas afir-

maciones no son falsas. Ocurre, más bien, que las frases en cuestión carecen de sentido. Esto no quiere decir que sean ridículas o tontas. Significa que a estas combinaciones de palabras no se les ha asignado ningún sentido y, en consecuencia, no dicen nada en absoluto, aunque parezca que sí lo hagan.

Réplica a la tercera objeción (Blakemore) de que la aplicación de predicados psicológicos al cerebro es meramente metafórica

La tercera objeción metodológica la planteó Colin Blakemore. Sobre la observación de Wittgenstein de que «sólo del ser humano y de lo que se parece a un ser humano (se comporta como tal) se puede decir: tiene sensaciones; ve, es ciego; oye, es sordo; es consciente o inconsciente», Blakemore señala que «resulta trivial, quizá simplemente errónea». Respecto a la acusación de que el aserto de los neurocientíficos de que en el cerebro hay «mapas» está preñado de posibilidades de confusión (ya que todo lo que se puede decir es que es posible mapear, por ejemplo, aspectos de los elementos del campo visual sobre la activación de las células de la corteza estriada «visual»), Blakemore señala que existen pruebas abrumadoras de «patrones topográficos de actividad» en el cerebro.

> Desde los tiempos de Hughlings Jackson, el concepto de subdivisión funcional y representación topográfica se ha convertido en imprescindible para la investigación sobre el cerebro. La tarea de cartografiar el cerebro está lejos de haberse completado, pero los éxitos del pasado nos hacen confiar en que, probablemente, cada parte del cerebro (y en especial la corteza cerebral) *esté* organizada topográfica-

mente. Como ocurre al descifrar una clave, en la traducción del sistema de escritura lineal B, o en la lectura de los jeroglíficos, todo lo que necesitamos para reconocer el orden del cerebro es un conjunto de reglas, reglas que relacionen la actividad de los nervios con acontecimientos que se producen en el mundo exterior o en el cuerpo del animal.[30]

El término «representación», por cierto, sólo significa aquí conectividad causal sistemática. Es algo inocuo. Pero no se debe confundir con el sentido en que se puede decir que una frase de una lengua representa el estado de cosas que describe, que un mapa representa aquello de lo que es un mapa, o que una pintura representa aquello de lo que es una pintura. No obstante, tal ambigüedad en el uso de «representación» es peligrosa, ya que es previsible que lleve a una confusión de los diferentes sentidos. Las siguientes observaciones de Blakemore demuestran cuán confusa puede ser:

Vistas estas pruebas tan abrumadoras de los patrones topográficos de la actividad del cerebro, no cabe extrañarse de que los neurofisiólogos y los neuroanatomistas hayan llegado a decir que el cerebro tiene *mapas*, que supuestamente desempeñan un papel esencial en la representación y la interpretación que el cerebro hace del mundo, del mismo modo que los mapas de un atlas lo hacen para sus lectores. El biólogo J. Z. Young describe el cerebro como dotado de un lenguaje de tipo pictográfico: «Lo que ocurre en el cerebro debe proporcionar una representación fiel de los acontecimientos que se producen fuera de él, y la disposición de las células en su interior ofrece un modelo detallado del mundo. El cerebro comunica significados mediante analogías topográficas».[31] ¿Pero existe algún peligro en el uso metafórico de términos como «lenguaje», «gramática» y «mapa» para describir las propiedades del cerebro?... No puedo imaginar que algún neurofi-

> siólogo crea que existe un cartógrafo fantasmagórico que esté hojeando el atlas cerebral. Tampoco pienso que el uso de palabras del lenguaje común (como mapa, representación, código, información e incluso lenguaje) sea un craso error conceptual de ese tipo [supuesto]. Tal imaginería metafórica es una mezcla de descripción empírica, licencia poética y vocabulario inadecuado.[32]

La existencia de cualquier posible peligro en el uso metafórico de las palabras depende de lo claro que esté que es meramente metafórico, y de si el autor tiene en mente que esto *es* todo lo que hay. Es muy dudoso que las adscripciones que los neurocientíficos hacen al cerebro de atributos que, en su literalidad, sólo se pueden aplicar a un animal como un todo sean en verdad meramente metafóricas (metonimias o sinécdoques). Los neurofisiólogos no creen, *por supuesto*, que haya un «cartógrafo fantasmagórico» que se dedique a hojear un atlas cerebral, pero *sí* piensan que el cerebro *hace uso* de los mapas. Según Young, el cerebro *construye hipótesis*, y lo hace *sobre la base* de esta «representación organizada topográficamente».[33] La cuestión es: ¿qué conclusiones sacan los neurocientíficos de su afirmación de que en el cerebro hay mapas o representaciones, o de su tesis de que el cerebro contiene información, o de que hay (como sugiere J. Z. Young) «lenguajes del cerebro»? Estos usos supuestamente metafóricos son en gran medida trampas en el camino del propio usuario. Éste *no necesariamente* caerá en ellas, pero es muy probable que lo haga.

La confusión de Blakemore

En la anterior cita de Blakemore se pone de manifiesto lo fácil que es que lo que se supone que es una metáfora inocua conduzca a la con-

fusión. Porque aunque pueda ser inocuo hablar de «mapas», es decir, de *mapear* características del campo perceptivo sobre grupos de células topográficamente relacionadas que reaccionan de manera sistemática a esas características, nada tiene de inocuo decir que tales «mapas» desempeñan «un papel esencial en la *representación* y la *interpretación* que el cerebro hace del mundo, del mismo modo que los mapas de un atlas lo hacen para sus lectores» (la cursiva es nuestra). En primer lugar, no está claro qué sentido hay que dar al término «interpretación» en este contexto. Porque no es en modo alguno evidente lo que se quiere decir con la afirmación de que las relaciones topográficas entre grupos de células que están relacionadas sistemáticamente con características del campo perceptivo desempeñan un papel esencial en el hecho de que el cerebro *interprete* algo. Interpretar significa, en su sentido literal, explicar el significado de algo, o considerar que algo que resulta ambiguo tiene un significado y no otro. Pero no tiene sentido suponer que el cerebro *explica* algo, o que entiende que algo *significa* una cosa y no otra. Si nos fijamos en J. Z. Young para averiguar lo que tenía en mente, lo que encontramos es la afirmación de que el cerebro «construye hipótesis y programas» a partir de tales mapas, lo cual no hace sino hundirnos más en la ciénaga.

Y lo que es más importante, cualquiera que sea el sentido que podamos dar a la afirmación de Blakemore de que los «mapas del cerebro» (que en realidad no son mapas) desempeñan un papel esencial en la «representación y la interpretación» que el cerebro hace del mundo, *no puede* ser «*del mismo modo que los mapas de un atlas lo hacen para sus lectores*». Porque un mapa es una representación pictórica, hecha de acuerdo a convenciones sobre el trazado de mapas y a reglas de proyección. Quien sepa leer un atlas debe conocer y entender estas convenciones, y extraer, a partir de los mapas, las ca-

racterísticas de lo que se representa. Pero los «mapas» del cerebro no son en absoluto mapas en este sentido. El cerebro no es lo mismo que el lector de un mapa, ya que no se puede decir de él que conozca convención alguna sobre representaciones o métodos de proyección, ni que sea capaz de leer nada a partir de la disposición topográfica de la células en estado de activación y de acuerdo a un conjunto de convenciones. Y es que las células no están en absoluto dispuestas de acuerdo a convenciones, y la correlación entre su estado de activación y las características del campo perceptivo no es convencional, sino *causal*.[34]

PHILOSOPHICAL FOUNDATIONS OF NEUROSCIENCE

Fragmento del capítulo 10

MAXWELL BENNETT y PETER HACKER

10.3 LOS *QUALIA*

Los qualia *entendidos como el carácter cualitativo de la experiencia: la concepción de los filósofos*

La tentación de ampliar el concepto de conciencia hasta abarcar todo el dominio de la «experiencia» se vio reforzada con la equivocada introducción por parte de los filósofos de la idea de los *qualia*. Lamentablemente, los neurocientíficos adoptaron esta idea aberrante y los malentendidos asociados con ella. El término «*qualia*» se introdujo para referirse al supuesto «carácter cualitativo de la experiencia». Cada experiencia, se dice, tiene un carácter cualitativo distintivo. Los *qualia*, sostiene Ned Block, «incluyen lo que se siente al ver, oír y oler, lo que se siente al tener dolor; más en general, cómo es tener estados mentales. Los *qualia* son propiedades experienciales de las sensaciones, los sentimientos, las percepciones y [...] también de los pensamientos y los deseos».[1] Asimismo, Searle sostiene que «todo estado consciente lleva consigo cierto sentimiento cualitativo, y esto se puede comprobar si se consideran algunos ejemplos. La experiencia de saborear la cerveza es muy diferente de la de escuchar

la *Novena Sinfonía* de Beethoven, y ambas tienen un carácter cualitativo distinto del de oler una rosa o contemplar una puesta de sol. Estos ejemplos ilustran las diferentes características cualitativas de las experiencias conscientes».[2] Al igual que Block, Searle también defiende que el hecho de pensar lleva consigo una sensación cualitativa especial. «Hay algo que es como pensar que dos más dos son cuatro. No hay manera de describirlo, a no ser diciendo que es el carácter del pensar conscientemente que "dos más dos son cuatro"».[3] «La mejor forma de caracterizar» el objeto de una investigación de la conciencia, señala Chalmers, es «la cualidad subjetiva de la experiencia». Un estado mental, dice, es consciente «si tiene una *sensación cualitativa*, una cualidad de la experiencia asociada a él. Estas sensaciones cualitativas se conocen también como cualidades fenoménicas o, más brevemente, *qualia*. El problema de explicar estas cualidades fenoménicas no es más que el problema de explicar la conciencia».[4] También es de la opinión de que pensar es una experiencia con un contenido cualitativo: «Cuando pienso en un león, por ejemplo, parece que haya un tono de cualidad leonina en mi fenomenología: lo que se siente al pensar en un león es sutilmente distinto de lo que se siente al pensar en la torre Eiffel».[5]

Los neurocientíficos siguen a los filósofos

Los neurocientíficos han secundado la idea de los *qualia*. Ian Glynn sostiene que «aunque aquello con lo que con mayor evidencia se asocian los *qualia* son las sensaciones y las percepciones, también se encuentran en otros estados mentales, como las creencias, los deseos, las esperanzas y los temores, durante episodios conscientes de estos estados».[6] Damasio afirma que «los *qualia* son las simples cualidades

sensoriales que encontramos en el azul del cielo o en el tono del sonido de un violoncelo, y, por consiguiente, los componentes fundamentales de las imágenes [en los que supuestamente consiste la percepción] consisten en *qualia*».[7] Edelman y Tononi sostienen que «cada experiencia consciente diferenciable representa un *quale* distinto, tanto si es primariamente una sensación, una imagen, un pensamiento o incluso un humor [...]»,[8] para luego afirmar que «el problema de los *qualia*» es «tal vez el problema más desalentador de la conciencia».

Explicación del carácter cualitativo de la experiencia desde la perspectiva de que al tenerla se siente algo

La sensación subjetiva o cualitativa de una experiencia consciente se caracteriza, a su vez, desde la perspectiva de que *algo se siente* cuando un organismo tiene una experiencia. «Una experiencia u otra entidad mental subjetiva es "fenoménicamente consciente" —dice la *Routledge Encyclopaedia of Philosophy*— sólo si hay algo que uno siente al tenerla.»[9] «Los estados conscientes son cualitativos —explica Searle— en el sentido de que para todo estado consciente [...] hay algo que se siente cualitativamente al estar en ese estado.»[10] La idea, y el uso mistificador de la expresión «hay algo que se siente» o «hay algo que es como...», proceden de un artículo del filósofo Thomas Nagel titulado «¿Qué se siente al ser un murciélago?». Nagel sostiene que «el hecho de que un organismo tenga experiencias conscientes significa, básicamente, que hay algo que se siente al *ser* ese organismo [...] Fundamentalmente un organismo tiene estados mentales conscientes si y sólo si hay algo que se siente al *ser* ese organismo, si ser ese organismo es algo *para* el organismo».[11] Esto, es decir, lo que el orga-

nismo siente al ser ese organismo, es el carácter o la cualidad subjetivos de la experiencia.

Explicación que Nagel hace de la conciencia desde la perspectiva de que «hay algo que se siente como...»

Si damos por supuesto que entendemos la frase «hay algo que se siente como» así usada, entonces parece que la idea de Nagel nos aclara el concepto de una criatura consciente y el de una experiencia consciente:

1) *Una criatura es consciente o tiene experiencia si y sólo si para la criatura hay algo que se siente como ser la criatura que es.*
2) *Una experiencia es una experiencia consciente si y sólo si para el sujeto que la tiene hay algo que se siente como tener la experiencia.*

De modo que, para un murciélago, hay algo que se siente como ser el murciélago (aunque Nagel afirma que no podemos imaginar *qué* es), y hay algo que se siente como ser un ser humano (y afirma que todos sabemos qué sentimos al ser nosotros).

Es importante observar que la frase "hay algo *que se siente como* tener una experiencia E para un sujeto" *no* indica *una comparación.* Nagel no dice que tener una determinada experiencia consciente *se parezca* a algo (por ejemplo, a alguna otra experiencia), sino que hay algo que se siente como tenerla *para el sujeto,* es decir, «que se siente como...» se pretende que signifique «cómo se siente para el propio sujeto».[12] Sorprende, sin embargo, que Nagel nunca diga, respecto a ni siquiera una sola experiencia, cómo es el tenerla. Sostiene que el carácter cualitativo de las experiencias de otras especies puede estar

más allá de nuestra capacidad de concebir. De hecho, lo mismo se puede decir de las experiencias de otros seres humanos. «El carácter subjetivo de la experiencia de una persona sorda y ciega de nacimiento, por ejemplo, no me es accesible ni la mía le es accesible a ella.» Pero sabemos cómo es ser nosotros, «y aunque no poseamos el vocabulario para describirlo adecuadamente, su carácter subjetivo es altamente específico y, en algunos aspectos, descriptible en términos que sólo criaturas como nosotros podemos entender».[13]

Los filósofos y los neurocientíficos coinciden

Filósofos y neurocientíficos han secundado esta idea. Les parece que capta la naturaleza esencial de los seres conscientes y la experiencia consciente. Así, Davies y Humphries defienden que, «pese a que no hay nada que se sienta como ser un ladrillo, o una impresora de inyección de tinta, probablemente hay algo que se siente como ser un murciélago, o un delfín, y sin duda hay algo que se siente como ser un ser humano. Un sistema, sea una criatura o un artefacto, es consciente sólo en el caso de que haya algo que se sienta como ser ese sistema».[14] Edelman y Tononi convienen en que «ya sabemos qué se siente al ser nosotros mismos, pero queremos explicar por qué somos conscientes, por qué hay "algo" que es como ser nosotros, explicar, en fin, cómo se generan las cualidades subjetivas experienciales».[15] Y Glynn sostiene que, en lo que a nuestras experiencias se refiere, por ejemplo las de oler el café recién molido, escuchar las notas del oboe o contemplar el azul del cielo, «sabemos cómo se siente al tener estas experiencias sólo porque las tenemos o las hemos tenido. [...] Del mismo modo que al oler el café recién molido uno se siente de una determinada forma, también uno se puede sentir de una determi-

nada forma (al menos intermitentemente) al creer que... o al desear que... o al temer que...».

Así pues, los *qualia* se conciben como las características cualitativas de los «estados mentales» o las «experiencias», categorías que incluyen no sólo la percepción, la sensación y el afecto, sino también el deseo, el pensamiento y la creencia. Para toda «experiencia consciente» o cada «estado mental consciente», hay algo que se siente como tenerla o estar en él. Este algo es un *quale*: un «sentimiento cualitativo». «El problema de explicar estas cualidades fenoménicas —dice Chalmers— no es otro que el problema de explicar la conciencia.»[16]

10.31 «Qué se siente» al tener una experiencia

Principales razones para extender el concepto común de conciencia

Una de las razones ofrecidas para extender el concepto de conciencia más allá de sus prudentes y legítimas fronteras es que lo que las experiencias tienen de distintivo, notable y, sin duda, misterioso es *que hay algo que se siente como tenerlas* . Se argumenta que una experiencia *es* una experiencia *consciente* sólo en el caso de que para el sujeto de la experiencia haya algo «que se siente como» tenerla. La conciencia, así concebida, se *define* en términos de *la sensación cualitativa de la experiencia*. Hay una forma específica de sentir al ver, oír y oler, sentir un dolor, o incluso al «tener estados mentales» (Block); todo estado consciente conlleva una determinada *sensación cualitativa* (Searle), y cada experiencia consciente distinguible representa un *quale* distinto (Edelman y Tononi). Esta sensación cualitativa, exclusiva de

toda experiencia distinguible, es *lo que se siente como tener la experiencia para el sujeto de la experiencia*. O esto es lo que se sostiene.

El uso de esas extrañas frases para invocar algo que se supone que a todos nos es familiar debería levantar nuestras sospechas. En primer lugar, analizaremos las «formas de sentir» y, después, ese «algo que se siente como...».

¿Hay siempre una manera específica de sentir el tener una «experiencia consciente»?

¿En verdad hay un *modo específico* en que se siente el ver, oír u oler? De hecho, podemos preguntarle a alguien que haya recuperado la vista, el oído o el sentido del olfato «¿qué se siente al ver (oír, oler) de nuevo?». Cabe esperar que la persona responda: «Es maravilloso», o quizá: «Es una sensación muy extraña». La pregunta se refiere a la actitud de la persona ante su experiencia y su recuperada capacidad de percepción, por ejemplo que le parece fantástico poder ver de nuevo, o extraño volver a oír después de tantos años de sordera. En estos casos, hay un modo en que se siente el ver u oír de nuevo, concretamente es fantástico o extraño. Pero si interrogásemos a una persona corriente sobre qué se siente al ver la mesa, la silla, el pupitre, la alfombra, etc., se preguntaría qué es lo que nos proponemos. Nada hay de distintivo en el hecho de ver esos objetos cotidianos. Evidentemente, ver la mesa difiere de ver la silla, el pupitre, la alfombra, etc., pero la diferencia no consiste en el hecho de que ver el pupitre *se siente diferente* de ver la silla. Ver una mesa o una silla corrientes no provoca *ninguna* reacción emocional ni actitudinal en circunstancias normales. Las experiencias difieren en la medida en que difieren sus objetos.

Podríamos decir, torpemente, que hay un modo de sentir un dolor. No es más que una manera enrevesada de decir que existe una respuesta a la pregunta (más bien tonta) «¿Qué se siente al tener un dolor?», por ejemplo, que es muy desagradable o, en algunos casos, terrible. De modo que se podría decir que cuando se sufre de migraña se siente algo distintivo, concretamente algo muy desagradable. Esta afirmación es inocua, pero no aporta ninguna prueba de la tesis general de que, para cada experiencia distinguible, hay una manera específica de sentir el tenerla. Los dolores son una excepción ya que, por definición, tienen un tono hedónico negativo. Los dolores son sensaciones intrínsecamente desagradables. Sin embargo, percibir no es una cuestión de tener sensaciones. Y la percepción en sus diversas modalidades y con sus indefinidamente numerosos objetos posibles, no es normalmente, aunque pueda serlo en ocasiones, el sujeto de cualidades afectivas o actitudinales cualesquiera (por ejemplo, agradable, placentera, horrible). Ni mucho menos el sujeto de una cualidad diferente para cada objeto de cada modalidad perceptiva. Además, para una gran diversidad de cosas que se pueden llamar «experiencias», no hay «una manera específica de sentir» el tenerlas, es decir, no existe respuesta a la pregunta «¿Qué se siente al...?».

Uno no puede sino estar de acuerdo con Searle en que la experiencia de saborear la cerveza es muy distinta de la de escuchar la *Novena* de Beethoven, y en que ambas son diferentes de las de oler una rosa o contemplar una puesta de sol. Puesto que las experiencias perceptivas se identifican o especifican esencialmente por su modalidad, es decir, la vista, el oído, el gusto, el olor y el tacto, y por sus objetos, es decir, aquello de que son experiencias. Pero sostener que las diversas experiencias tienen una *sensación* exclusiva y distintiva es completamente diferente y mucho más cuestionable. Es más cues-

tionable en la medida en que no está claro lo que *significa*. Es evidente que las cuatro experiencias que Searle cita son normalmente agradables para mucha gente. Y es totalmente cierto que la identidad del placer o deleite depende del objeto del placer. No se puede obtener el placer de tomarse una cerveza con la audición de la *Novena* de Beethoven, ni el placer de contemplar una puesta de sol en el aroma de una rosa. Ésta es una verdad lógica, no empírica, es decir, no es que la sensación cualitativa distintiva de contemplar una puesta de sol difiera fácticamente de la sensación distintiva de oler una rosa —al fin y al cabo, ambas pueden ser muy agradables—. Lo que ocurre es que el placer de ver una puesta de sol difiere lógicamente del de oler una rosa, pues la identidad del placer depende de aquello que lo produce. De esto no se sigue que toda experiencia tenga un *carácter cualitativo* diferente, es decir, que haya una sensación específica para todas las experiencias. Porque, en primer lugar, la mayoría de las experiencias no tienen, en este sentido, carácter cualitativo alguno; ni producen placer ni displacer, no son ni agradables ni desagradables, etc. Mientras andamos por la calle, vemos cientos de objetos diferentes. Ver una farola es una experiencia distinta de la de ver un buzón, pero ¿se siente algo distinto en cada caso? No; ni tampoco se siente lo mismo, ya que la visión de ambos objetos no provoca respuesta alguna, no hay ningún tipo de sentimiento cualitativo asociado a la visión de ninguno de ellos. En segundo lugar, diferentes experiencias que *sí* tienen asociada una sensación cualitativa, por ejemplo con un componente hedónico, pueden tener asociada la misma sensación. Lo que las diferencia no es cómo se sienten, pese a que la pregunta «¿Qué se siente al V?» (donde «V» especifica una experiencia apropiada) puede tener exactamente la misma respuesta, ya que diferentes experiencias pueden ser igualmente placenteras o desagradables, interesantes o aburridas.

El carácter cualitativo de las experiencias correctamente concebidas

Tanto sentir un dolor como percibir lo que sea que uno percibe se pueden llamar «experiencias». También se puede llamar así el hecho de estar en un determinado estado emocional. Y también, por supuesto, realizar una variedad indefinida de actividades. Podríamos decir que las experiencias son los sujetos posibles de predicados actitudinales, es decir, pueden ser agradables o desagradables, interesantes o aburridas, estupendas o terribles. Es a estos atributos a los que se les puede llamar «caracteres cualitativos de las experiencias», no a las propias experiencias. Por lo tanto, no se puede decir de modo inteligible que ver el color rojo o contemplar el *Guernica*, oír un sonido o escuchar *Tosca*, son «*qualia*». En consecuencia, cuando Damasio dice que el carácter azul del cielo es un *quale*, está cambiando el sentido del término «*quale*», ya que si el color de un objeto es un *quale*, entonces los *qualia* no son en modo alguno las características cualitativas de las experiencias, sino las cualidades de los objetos de la experiencia (o, si se considera que los colores no son cualidades de los objetos, entonces los *quale* son los constituyentes del contenido de las experiencias perceptivas). Asimismo, cuando Edelman y Tononi afirman que cada experiencia consciente distinguible representa un *quale* distinto, sea una sensación, una imagen, un estado de ánimo o un pensamiento, están cambiando el sentido del término «*quale*». Porque es evidente que «*quale*» no significa «el carácter cualitativo de una experiencia» en el sentido que hemos examinado. Lo que sí significa, o debería significar, es algo que analizaremos en breve (§ 10.34).

Conviene señalar que decir que una experiencia es un sujeto de un predicado actitudinal es una forma de hablar potencialmente

engañosa. Porque decir que una experiencia (por ejemplo, ver, observar, vislumbrar, escuchar, saborear esto o lo otro, pero también andar, charlar, bailar, jugar, escalar montañas, combatir en batallas, pintar cuadros) lleva aparejada una determinada sensación (por ejemplo, que es agradable, deliciosa, encantadora, desagradable, repugnante, asquerosa) no es más que decir que el sujeto de la experiencia, es decir, la persona que vio, escuchó, saboreó, anduvo, charló, bailó, etc., encontró agradable, delicioso, encantador etc. hacerlo. Por lo tanto, el carácter cualitativo de una experiencia E, esto es, qué se siente al tener esa experiencia, es la actitud afectiva del sujeto (cómo se sintió) al experimentar E.

Para evitar confundirse en este punto, debemos distinguir cuatro cuestiones:

1) Muchas experiencias se individualizan esencialmente, es decir, se distinguen, especificando aquello de que son experiencia.

2) Toda experiencia es un sujeto *posible* de predicados actitudinales positivos y negativos, por ejemplo, predicados de placer, interés o atracción. De ahí no se sigue, y además es falso, que toda experiencia sea un sujeto *real* de un predicado actitudinal positivo o negativo.

3) Es posible que distintas experiencias, cada una de ellas sujeto de un atributo actitudinal, no se puedan distinguir por cómo se siente la persona que las experimenta. El olor de las rosas es distinto del de las lilas. Oler rosas y oler lilas son experiencias diferentes. Uno no puede obtener el placer de oler una rosa oliendo para ello una lila. Pero bien puede ser que las experiencias sean agradables por igual. De modo que, si se nos pregunta qué se siente al oler una rosa y qué se siente al oler una lila, la respuesta puede ser perfectamente la misma, a saber, «algo delicioso». Si esta respuesta especifica

cuál fue la sensación, entonces es obviamente falso que toda experiencia se pueda individualizar de forma exclusiva por su carácter cualitativo distintivo o *quale*. No debemos confundir el carácter cualitativo de la experiencia con el carácter cualitativo del objeto de la experiencia. Es esto último, y no lo primero, lo que individualiza la experiencia.

4) Incluso si ampliamos el concepto de experiencia para que incluya el pensar que algo es como es, o el pensar en algo, lo que diferencia esencialmente el pensar una cosa y no otra no es qué se siente o cómo se siente al pensar lo que se piensa. Pensar que $2 + 2 = 4$ es distinto de pensar que $25 \times 25 = 625$, y ambas cosas difieren de pensar que los demócratas ganarán las próximas elecciones.[17] Difieren en la medida en que son sus objetos los que las especifican o individualizan esencialmente. Podemos pensar *que* algo es así o asá o pensar *en* una u otra cosa sin tener a la vez ningún tipo de actitud afectiva; luego no es necesario que haya ningún «qué se siente» al pensar eso. Pensar en leones, en Ricardo Corazón de León o en las calles de León, puede ir acompañado de un hálito leonino. Pero, contrariamente a lo que propone Chalmers, especificar el hálito asociado no es caracterizar *qué se siente* al pensar en esas cosas, y mucho menos caracterizar el pensamiento de forma exclusiva para individualizarlo. El hecho de que alguien asocie pensar en una de esas cosas con un hálito leonino no responde a la (extraña) pregunta «¿Qué se siente al pensar en leones (Ricardo Corazón de León, las calles de León)?», y desde luego no distingue el pensar en leones del pensar en León o en Ricardo I.

PHILOSOPHICAL FOUNDATIONS OF NEUROSCIENCE

Fragmento del capítulo 14 Observaciones finales

MAXWELL BENNETT y PETER HACKER

14.5 POR QUÉ IMPORTA

Sobre la cuestión de cómo afectará al próximo experimento

Imaginemos a un científico que lee nuestras discusiones analíticas con cierta perplejidad. Es posible que esté moderadamente interesado en alguno de nuestros análisis de conjunto, pero a la vez se siente desconcertado ante lo que parece ser un interminable desmenuzamiento lógico. Tras leer nuestras exposiciones iniciales, tal vez se pregunte: «¿Es todo esto realmente importante?». «Después de todo —es posible que se diga—, ¿en qué afectará al próximo experimento?» Confiamos en que todo lector que nos haya seguido hasta aquí no se sienta tentado de hacerse esa pregunta, porque denota incomprensión.

El hecho de que las reflexiones analíticas afecten o no a los próximos experimentos *no* nos preocupa. Es posible que lo hagan o no —depende del experimento en cuestión, y de cuáles sean los supuestos del neurocientífico—. Por lo dicho hasta aquí, debería resultar obvio que, si nuestros argumentos son convincentes, lo mejor que se puede hacer con algunos experimentos es abandonarlos.[1] Otros se de-

berían diseñar de nuevo.[2] Es posible que a la mayoría no los afecte, aunque probablemente deban reformularse las hipótesis de partida y describir los resultados de forma muy distinta.[3]

Lo que nos interesa es comprender el último experimento

Nuestro interés no radica en el diseño del próximo experimento, sino en la *comprensión* del último. De modo más general, las investigaciones conceptuales contribuyen principalmente a la comprensión de lo que se sabe y a la claridad en la formulación de preguntas referentes a lo que no se sabe. No tendría la menor importancia que nuestras reflexiones *no* tuvieran efecto en el experimento siguiente, pero sí tienen un efecto considerable en la interpretación de los resultados de los experimentos anteriores. Y no hay duda de que tienen algo que aportar al planteamiento de preguntas, a su formulación y a la distinción entre preguntas significativas y preguntas confusas. (Si estamos en lo cierto, entonces las preguntas sobre el «problema de la integración», entendido como el problema de cómo el cerebro forma imágenes, son en gran medida expresiones de confusión,[4] y gran parte del debate sobre imaginería mental se basa en un malentendido.)[5]

¿Tiene alguna importancia? Sí, si la tiene la comprensión

¿Es que todo este aparente desmenuzamiento lógico, todo este detallado examen de las palabras y de su uso, son *importantes*? ¿Es que la neurociencia realmente *necesita* todo esto? Si el espíritu impulsor tras el empeño neurocientífico es el deseo de entender los fenómenos neurales y su relación con las capacidades psicológicas y su ejercicio, en-

tonces importa muchísimo. Y es que, con independencia de la brillantez de los experimentos del neurocientífico y del refinamiento de sus técnicas, si existen confusión conceptual en sus preguntas o errores conceptuales en la descripción de los resultados de sus investigaciones, es que no ha entendido aquello que se proponía entender.

Muchos neurocientíficos contemporáneos que trabajan en el ámbito de la neurociencia cognitiva coinciden en que la defensa por parte de sir John Eccles de una forma de dualismo fue un error[6] —y lo que está en la base del error de Eccles es una *confusión conceptual*—. Mediante la referencia a diversas teorías de distinguidos neurocientíficos cognitivos contemporáneos hemos intentado demostrar que, lejos de haber sido erradicado con el rechazo superficial de las diversas formas del dualismo cartesiano, el error conceptual sigue muy extendido. Afecta e infecta a la plausibilidad de las preguntas planteadas, al carácter de los experimentos ideados para responderlas, a la inteligibilidad de las descripciones de los resultados de estos experimentos y a la coherencia de las conclusiones que de ellos se derivan. Y esto es sin duda importante, tanto para la comprensión de lo que los actuales neurocientíficos *han* conseguido como para el progreso de la neurociencia cognitiva en el futuro.

¿Por qué es importante para el público culto?

También es muy importante para el público culto porque, al margen de que ciertos neurocientíficos estén o no confundidos, no hay duda de que las formas de descripción que emplean sí confunden al público no especializado. Es comprensible que los neurocientíficos estén ansiosos por comunicar los conocimientos sobre el funcionamiento del cerebro que han alcanzado durante las últimas décadas, y por

compartir con el público culto la pasión que sienten por su objeto de estudio. Así lo demuestra la avalancha de libros escritos por numerosos miembros distinguidos de la profesión. Pero cuando hablan del pensamiento y el razonamiento del cerebro, de que uno de sus hemisferios sabe algo de lo que no informa al otro, de que el cerebro toma decisiones sin que la persona lo sepa, de imágenes mentales que rotan en el espacio mental, etc., los neurocientíficos están fomentando una forma de mistificación, cultivando una neuromitología deplorable en todos los sentidos. Porque, en primer lugar, con ello lo hacen todo menos generar comprensión en el público no especializado al que se dirigen. En segundo lugar, el público no especializado buscará en la neurociencia respuestas a seudopreguntas que no debería formular y que la neurociencia no puede responder. Y en adelante ese público, decepcionado, ignorará las preguntas realmente importantes que la neurociencia *puede* tanto formular como responder. Y esto es, sin duda, importante.

Sobre la necesidad de claridad conceptual

A lo largo de este libro, hemos intentado demostrar que la claridad referente a las estructuras conceptuales es tan importante para la neurociencia cognitiva como la claridad sobre los métodos experimentales. La claridad conceptual ilumina, en vez de obstaculizar, las grandes aportaciones de la neurociencia cognitiva a nuestra comprensión de las raíces biológicas de las capacidades humanas. Y es que los logros de la neurociencia sólo se pueden apreciar en su justa medida cuando se disipan las sombras causadas por las confusiones conceptuales.

NEUROCIENCIA Y FILOSOFÍA

MAXWELL BENNETT

UNA ODISEA PERSONAL

Cuando un potencial de acción en propagación alcanza una sinapsis en el extremo de una terminal axónica de una neurona presináptica, simbolizada por el pequeño triángulo en la figura 1, induce la liberación de moléculas neurotransmisoras, como se muestra en el recuadro de una sinapsis en el extremo inferior izquierdo de la figura 1. El transmisor se difunde a lo ancho de una estrecha hendidura y se une a los receptores de la membrana postsináptica. Tal unión induce la abertura de canales y, a menudo, la generación de potenciales de acción en la neurona postsináptica. Para este proceso se requieren varios cientos de proteínas (Sieburth y otros, 2005). He dedicado más de cuarenta años a la investigación de los mecanismos que intervienen en la transmisión en la sinapsis (Bennett, 2001), y hace poco inicié una serie de estudios sobre cómo las redes de sinapsis llevan a cabo sus funciones en el cerebro. Estas redes, formadas por miles de millones de neuronas, cada una con hasta unas diez mil sinapsis, se pueden encontrar en las partes del cerebro que deben funcionar normalmente para que se pueda recordar un acontecimiento nuevo durante más de un minuto (el hipocampo), para ver (la retina y la cor-

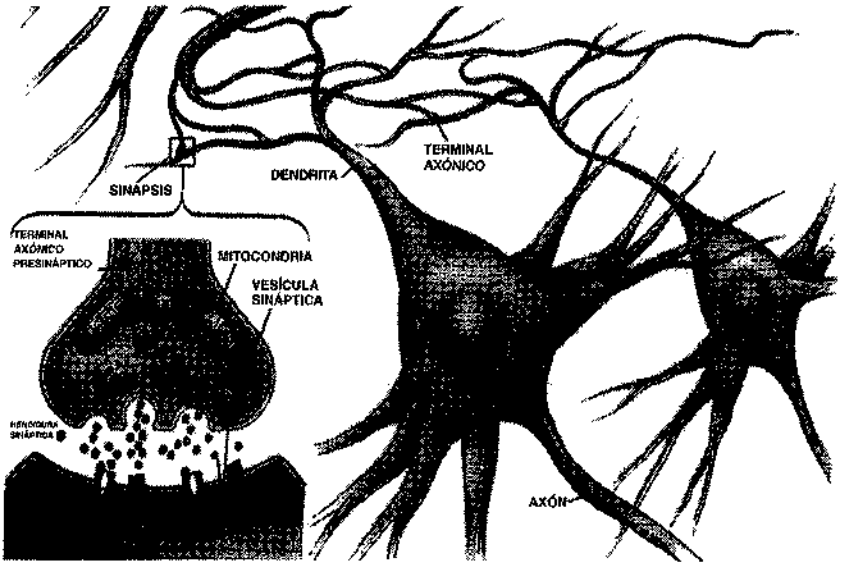

FIGURA 1: En los axones de dos neuronas tienen lugar procesos que culminan en el íntimo contacto entre ellas en puntos llamados sinapsis. En el recuadro aparece una sinapsis, ampliada en la esquina izquierda. En el texto se explica el funcionamiento de esta sinapsis.

teza visual primaria) y para adquirir una diversidad de habilidades motrices (cerebelo).

Las redes de sinapsis que escogí para el estudio inicial fueron las del hipocampo (Bennett, Gibson y Robinson, 1994). La distribución general de los tipos de neuronas del hipocampo y sus sinapsis, descrita por primera vez por Ramón y Cajal (1904), se muestran en la figura 2. Las letras que aparecen en la figura se refieren a las diferentes partes del hipocampo, así como a los tipos de neuronas (indicados por elipsoides negros, con axones y dendritas largos y estrechos), junto con sus conexiones sinápticas, como se señala en el texto de la figura. Un enfoque ingenieril de la descripción del funcionamiento del hipocampo requiere el desarrollo de una representación de una red neural como la que se muestra a la derecha de la figura 2 y se describe en

el texto de ésta. Los tipos de neuronas se indican aquí mediante círculos; sus dendritas y axones, mediante líneas rectas; y las sinapsis, mediante pequeños rectángulos. En mi juventud, recién obtenido el título de ingeniero eléctrico, me intrigaba la sugerencia de Brindley (1967) de que algunas de las sinapsis de redes como ésas eran modificables. Lo que quería decir con este término es que las sinapsis son permanentemente susceptibles de modificar sus propiedades como consecuencia de la llegada de un potencial de acción al terminal axónico. Entonces, dada esta posibilidad, podría ser que «mecanismos de memoria y condicionamiento del sistema nervioso almacenen información mediante sinapsis modificables». Más tarde el mismo Brindley (1969) demostró en modelos de redes neurales cómo tales sinapsis modificables podían «realizar muchas tareas de aprendizaje simples». Posteriormente, Marr (1971), que también trabajaba en la Universidad de Cambridge, señaló que «la característica más importante de la archicorteza (hipocampo) es su capacidad de realizar un tipo simple de tarea de memorización». Fue él quien sugirió por primera vez que una red recurrente de colaterales (fig. 2) podría actuar como memoria asociativa si la eficacia de las sinapsis excitadoras fuera modificable, y si los potenciales de membrana de las neuronas piramidales fueran establecidos por interneuronas inhibitorias que computan la actividad total de la red. Su sugerencia se formulaba en unos términos de ingeniería que a mí y a muchos otros nos resultaron muy atractivos para posteriores estudios teóricos y experimentales. Mis colegas y yo seguimos este enfoque conceptual general de Brindley y Marr, e identificamos las condiciones en las que podría funcionar una representación de red neural del hipocampo como la de la figura 2 (Bennett, Gibson y Robinson, 1994). Sugerimos que «la recuperación de un recuerdo empieza con la activación de un conjunto de neuronas piramidales que se solapa con el recuerdo que se

va a recuperar» y que «la activación de diferentes conjuntos de neuronas piramidales tiene lugar a continuación en pasos sincrónicos discretos» hasta que se recupera el patrón neuronal de memoria almacenado (fig. 2).[1] Sin embargo, hay dos aspectos de esta clase de enfoque ingenieril de la descripción del funcionamiento de las redes sinápticas y, por consiguiente del cerebro, que me parecen equívocos y que se detallan en los apartados que siguen.[2]

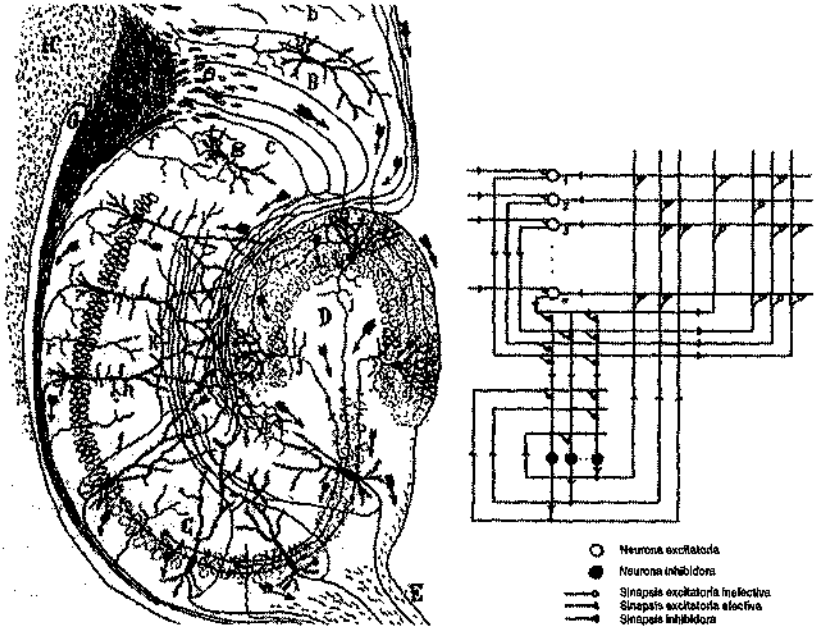

FIGURA 2: El dibujo de la izquierda representa las neuronas del hipotálamo y sus conexiones sinápticas tal como las plasmó en tinta Ramón y Cajal (Figura 479, en Ramón y Cajal, 1904). Las letras de la figura se refieren a los distintos componentes del hipocampo relevantes para nuestros propósitos: D, circunvolución dentada; K, colaterales recurrentes de las células piramidales de la región CA_3 del hipocampo (C), que permiten a éstas formar sinapsis entre sí además de proyectarse para formar sinapsis con las neuronas piramidales de la región CA_1 del hipocampo (h).

A la derecha hay un diagrama de la red sináptica básica de la región CA_3 del hipocampo, compuesta de neuronas piramidales (círculos vacíos) e interneuronas inhibidoras (círculos rellenos). Las neuronas piramidales establecen conexiones aleatorias entre sí a través de sus colaterales recurrentes. Antes del aprendizaje, estas conexiones son inefi-

(Continúa)

PRIMERA FUENTE PRINCIPAL DE INTERÉS RELATIVA A NUESTRA COMPRENSIÓN DE LA FUNCIÓN DE LAS REDES CELULARES DE LA CORTEZA CEREBRAL

Dado que sabemos que una lesión en el hipocampo impide que uno recuerde un suceso durante más de un minuto aproximadamente, ¿cómo se identifican las células del hipocampo normal que intervienen en nuestra capacidad de recordar, cuáles son las conexiones sinápticas entre estas células y cómo funcionan estas sinapsis? Por ejemplo, en cualquier volumen dado del hipocampo hay muchas

caces; después del aprendizaje, un subconjunto de ellas se hace eficaz y en el estado final de desarrollo de la red hay conexiones sinápticas excitadoras cuyas fuerzas se considera que tienen el valor de la unidad (triángulos vacíos), y otras cuyas fuerzas se han quedado en el valor cero (círculos vacíos). Las interneuronas inhibidoras reciben conexiones aleatorias desde muchas células piramidales y también desde neuronas inhibidoras. Las neuronas inhibidoras, a su vez, se proyectan hacia neuronas piramidales. Se considera que la fuerza de cualquier sinapsis en la que intervengan interneuronas inhibidoras es fija. El estado inicial del sistema lo determina un patrón de activación que llega a las neuronas piramidales desde los axones de fibra musgosa que surgen de la región D en el dibujo de Ramón y Cajal, o de la vía perforante directa que se encuentra encima de la región D, patrón representado por las líneas que entran desde la izquierda. Una vez establecido el estado inicial, la fuente externa se elimina y la red recurrente de CA_3 actualiza su estado interior cíclica y sincrónicamente.

En el modelo de red neural, las neuronas inhibidoras se representan como dispositivos lineales de acción rápida que producen outputs proporcionales a sus inputs, y que desempeñan una importante función reguladora en el establecimiento de los potenciales de membrana de las neuronas piramidales. La probabilidad de que una neurona se excite en un denominado recuerdo almacenado, que determina la cantidad media de neuronas activas cuando «se recupera un recuerdo», se puede fijar a voluntad. Se supone que los «recuerdos» de esta red se almacenan en las sinapsis colaterales recurrentes utilizando un hebbiano de dos valores. En la teoría se tienen en cuenta tanto las correlaciones espaciales entre las fuerzas aprendidas de las sinapsis colaterales recurrentes como las correlaciones temporales entre el estado de la red y estas fuerzas sinápticas. Se supone que la recuperación de un recuerdo empieza con la activación de un conjunto de neuronas piramidales de CA_3 que se superponen con el recuerdo a recuperar, así como con la activación de un conjunto de neuronas piramidales que no están en el recuerdo a recuperar; la activación de ambos conjuntos de neuronas probablemente es inducida por las sinapsis que los axones de la vía perforante forman sobre las neuronas de CA_3. La activación de diferentes conjuntos de neuronas piramidales se desarrolla a continuación en pasos sincrónicos discretos (para más detalles, véase Bennett, Gibson y Robinson, 1994).

más células gliales que neuronas, y estas células gliales son de diversos tipos, como ocurre con las neuronas. El descubrimiento de ondas que propagan y transmiten la actividad entre las células gliales astrocíticas (Cornell-Bell y otros, 1990), aunque de forma mucho más lenta que entre las neuronas (Bennett, Farnell y Gibson, 2005), introduce una considerable complejidad en la búsqueda de correlatos celulares de la memoria en el hipocampo. Aunque esas ondas gliales se han considerado irrelevantes para la búsqueda de correlatos celulares de nuestros atributos psicológicos (Koch, 2004), no se han realizado experimentos que lo demuestren fehacientemente. Dada la intimidad física entre las células gliales y las neuronas, será difícil averiguar si las primeras son relevantes o no en la búsqueda de correlatos celulares, a pesar de que podamos manipular genéticamente las proteínas de estos diferentes tipos de células. Creo que sería mucho más sensato buscar correlatos «celulares», y no simplemente «neuronales». Además, apenas estamos empezando a comprender la gran variedad de relaciones sinápticas entre las neuronas, entre las células gliales, y entre ambas clases de células, así como la diversidad de mecanismos sinápticos especializados que operan entre todas ellas. En mi opinión, está fuera de lugar el actual endiosamiento de las neurociencias, manifiesto, por ejemplo, en la presunción de que redes como las de la figura 2 proporcionan indicios importantes sobre el funcionamiento de las redes sinápticas de cualquier parte del cerebro. Con el fin de aportar evidencia a favor de este punto, expongo a continuación algunos ejemplos de cuán dolorosamente lento es el progreso en la comprensión incluso de las redes sinápticas relativamente simples que operan en la retina, la corteza visual primaria y el cerebelo.

Las redes de la retina

La parte más simple y accesible del sistema nervioso central es la retina, que durante el desarrollo aparece primero como una evaginación del cerebro. Por esta razón, el más importante histólogo del sistema nervioso, Ramón y Cajal, consideraba la retina el lugar ideal donde iniciar la investigación relativa a la comprensión del funcionamiento de las redes sinápticas del sistema nervioso central. Poco después de que yo iniciara la investigación, Barlow y Levick (1965) descubrieron las denominadas neuronas ganglionares direccionalmente selectivas en la retina de algunas especies. Una neurona ganglionar direccionalmente selectiva es aquella que envía impulsos a gran velocidad cuando un objeto se mueve en un determinado sentido (llamado, por consiguiente, sentido preferente) por delante de los fotorreceptores bastón sensibles a la luz que están conectados a la célula ganglionar, situados por encima de ella. Cuando el objeto se mueve en el sentido contrario, la neurona ganglionar no se activa (y, por consiguiente, a ese sentido se le llama sentido nulo). Barlow y Levick propusieron el esquema que se muestra en la figura 3 (parte inferior) para explicar los orígenes reticulares de la selectividad direccional. En esta red, una señal inhibidora compensatoria veta la señal excitadora de movimiento en el sentido nulo. Una interneurona lateral transmite una señal inhibidora en el sentido nulo, pero no en el preferente, mientras la señal excitadora actúa localmente (fig. 3). Por lo tanto, en el movimiento en el sentido nulo la inhibición llega antes que la excitación y puede interactuar con ella, pero en el movimiento en el sentido preferente la inhibición llega más tarde que la excitación. Este algoritmo se explica en el texto de la figura 3. Esta red tan simple me impresionó mucho cuando se publicó en 1967. Era el primer análisis de este tipo que daba una explicación del funcionamiento de una red sináptica real.

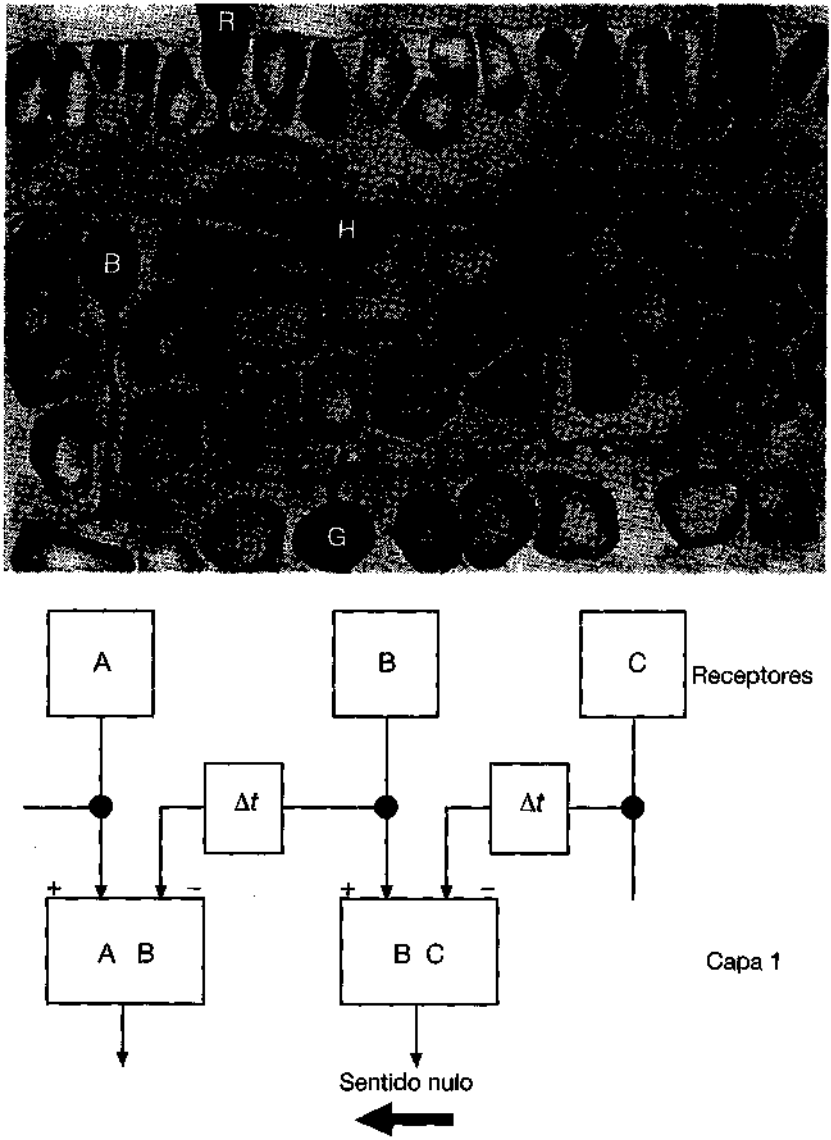

FIGURA 3: El gráfico superior es un diagrama de algunas de las células principales de la retina identificadas en 1965. R, fotorreceptor bastón; B, neurona bipolar; H, neurona horizontal; G, neurona ganglionar que conecta la retina con el cerebro.

El gráfico inferior muestra el algoritmo propuesto por Barlow y Levick (1965) para explicar la selectividad direccional de las neuronas ganglionares. A, B y C son receptores que pueden responder al objeto, que se mueve por encima de ellos sea en el sentido nulo

(Continúa)

Si este algoritmo es correcto, entonces se plantea la cuestión de la identidad de la interneurona lateral y sus conexiones sinápticas. ¿Hasta dónde hemos avanzado los neurocientíficos durante los últimos cuarenta años, desde que se sugirió el algoritmo, en la identificación de los componentes celulares que ejecutan los cómputos necesarios y sus conexiones? En el momento de la investigación original sobre la selectividad direccional, se conocía la existencia de unos diez tipos diferentes de células de la retina vertebrada (Ramón y Cajal, 1904; Polyak, 1941). Hoy se reconocen al menos cincuenta tipos diferentes de células (Masland, 2001), sin contar los distintos tipos de células gliales que entran en contacto íntimo con las neuronas pero no transmiten potenciales de acción. Han sido necesarios cuarenta años de investigación para identificar algunos de los mecanismos celulares de la retina responsables de la selectividad direccional, y muchas preguntas importantes siguen sin respuesta (Fried, Munch y Werblin, 2005). Se han identificado algunas de las características especificadas en el esquema de Barlow y Levick (1965) en una neurona llamada célula amacrina estrellada (Fried, Munch y Werblin, 2002). Sin embargo, el esquema se ha tenido que modificar radicalmente al descubrirse que los inputs de las células ganglionares son ellos mismos direccionalmente selectivos (Vaney y Taylor, 2002).

(señalado con una flecha), sea en el sentido preferente. Cada uno de estos receptores puede excitar la actividad de las unidades que están inmediatamente debajo de ellos. Cada recuadro que contiene una Δt es una unidad que si es excitada por el receptor conectado con ella impedirá, después de un retraso de longitud Δt, la excitación de la unidad adyacente en el sentido nulo. Cuando un objeto se mueve en sentido nulo, la actividad eléctrica de un receptor excitado (por ejemplo, C) excita (+) una unidad de la capa situada inmediatamente debajo de ella, y al mismo tiempo inhibe (–) la unidad siguiente en el sentido nulo; a su vez cada receptor, concretamente C, B y A, lleva a cabo este proceso a medida que el objeto se mueve por encima de ellos. Las unidades de retraso (señaladas como Δt) determinan que el proceso inhibitorio detenga la actividad excitadora de A y B que circula a través de estas puertas si el movimiento es en sentido nulo, pero alcanza las puertas demasiado tarde para producir tal inhibición si el movimiento es en sentido preferente.

Así pues, parece que en la circuitería retinal hay varios niveles que determinan la selectividad direccional de las neuronas ganglionares, que probablemente implican la actividad de al menos cuatro redes sinápticas retinales diferentes y específicas aún por identificar (Fried, Munch y Werblin, 2005). Esto ilustra las dificultades para la comprensión de una propiedad de la red de incluso lo que se ha considerado la parte más simple y accesible del sistema nervioso central.

Las redes de la corteza visual primaria

La conferencia científica más apasionante a la que he asistido tuvo lugar hace más de treinta años, en el Simposio sobre la Sinapsis de Cold Spring Harbor en 1975. James Watson me invitó a exponer lo que habíamos descubierto sobre la plasticidad de las sinapsis entre los nervios y las células musculares, en especial nuestros descubrimientos sobre el desarrollo del punto de contacto inicial del terminal del nervio y el músculo, cómo este punto se convierte en enclave de terminales nerviosas en exceso, y cómo todas menos una de estas terminales son eliminadas a lo largo del proceso (Bennett y Pettigrew, 1976). Antes de llegar al simposio ignoraba que, tras mi exposición, Hubel, Wiesel y LeVay (1976) ofrecerían una magnífica explicación del desarrollo de las conexiones de las redes sinápticas en la corteza visual primaria (V_1), que subyacen a la formación de columnas de neuronas dominadas por las conexiones con un ojo o con el otro. En las primeras fases del desarrollo, estas neuronas tienen conexiones con ambos ojos, pero tras un proceso de eliminación sináptica uno u otro ojo acaba por ser el dominante en términos de conectividad. Hubel, Wiesel y LeVay demostraron de forma realmente espectacu-

lar la considerable plasticidad de este proceso, mostrando cómo si la visión se limita a un solo ojo durante el inicio del desarrollo, el otro ojo domina la conectividad sináptica. Durante un período crítico del desarrollo es posible invertir este resultado, si se restaura la visión en ambos ojos.

Quince años después de aquella memorable exposición, en el Simposio de Cold Spring Harbor de 1990, Wiesel y sus colegas Gilbert y Hirsch retomaron el tema de la plasticidad de las conexiones sinápticas en la corteza visual primaria (V_1). Sin embargo, esta vez desplazaron el énfasis al grado de plasticidad de la red sináptica de la corteza visual de los adultos. Su investigación, basada en técnicas de grabación electrofisiológica, sugería que a los pocos meses de una lesión retinal tiene lugar una reorganización a gran escala de las redes sinápticas del interior de la corteza (Gilbert, Hirsch y Wiesel, 1990). Este trabajo se interpretó en el sentido de que las redes sinápticas corticales adultas podían reajustarse tras la pérdida de una señal sensorial de entrada (Gilbert, 1998). Entender con claridad los mecanismos de plasticidad responsables de esos reajustes es de considerable importancia si pretendemos ayudar a quienes necesitan un tratamiento de rehabilitación adecuado.

Entre los dos y los seis meses posteriores a lesiones retinales binoculares que priven de su input normal a una zona del interior de la corteza visual primaria (V_1), se detecta actividad dirigida por estímulos recurrente hasta 5 mm dentro del área de la corteza privada de su input visual (Gilbert, 1998). Se cree que cambios más modestos de la topografía cortical, de un alcance de 1-2 mm, se producen de forma inmediata (de minutos a horas) después de tales lesiones (Gilbert y Wiesel, 1992). Quince años después de estos estudios sobre la plasticidad de la corteza visual adulta, Logothetis y sus colegas estudiaron las señales en la corteza visual primaria del macaco después de

pequeñas lesiones retinales binoculares, con el fin de determinar el alcance y el curso temporal de la reorganización de la corteza visual (V_1; Smirnakis y otros, 2005). Las lesiones retinales se provocaron mediante fotocoagulación por láser, de modo que se creara un escotoma o vacío perceptivo de 4-8 grados de diámetro en el campo visual. Estas lesiones privaban del input visual de cada ojo a parte de la corteza, un proceso que supuestamente maximiza la reorganización. La región de la corteza visual a la que se priva del input retinal se llama zona de proyección de la lesión. Se emplearon imágenes por resonancia magnética funcional (RMf) para detectar los cambios en la topografía cortical de la corteza visual del macaco después de estas lesiones retinales binoculares. En contraste con los estudios antes mencionados en los que se utilizó electrofisiología, el amplio campo de visión que ofrecen las imágenes por RMf demostró que, durante los 7,5 meses posteriores a las lesiones retinales, la corteza visual primaria no se aproxima a su capacidad normal de respuesta y su topografía no cambia (fig. 4). Aprovechando el amplio campo de visión que esas imágenes proporcionan, se pudieron colocar electrodos grabadores electrofisiológicos exactamente en la zona de proyección de la lesión. Estos electrodos confirmaron los resultados de las imágenes por RMf. Así pues, según las observaciones de Logothetis y sus colegas utilizando dos técnicas diferentes, la corteza visual primaria tiene un potencial limitado para la reorganización de las redes sinápticas, al menos hasta algunos meses después de la lesión en la retina.

Todavía no se conocen los detalles acerca de lo que pudo salir mal en los últimos quince años de investigaciones sobre el tema de la plasticidad cortical. Basta con decir que la complejidad que supone determinar las propiedades de este primer enclave de la corteza que recibe input de la retina es tal que requiere un gran cuidado, habilidad técnica, perspicacia teórica y determinación. Wiesel, pre-

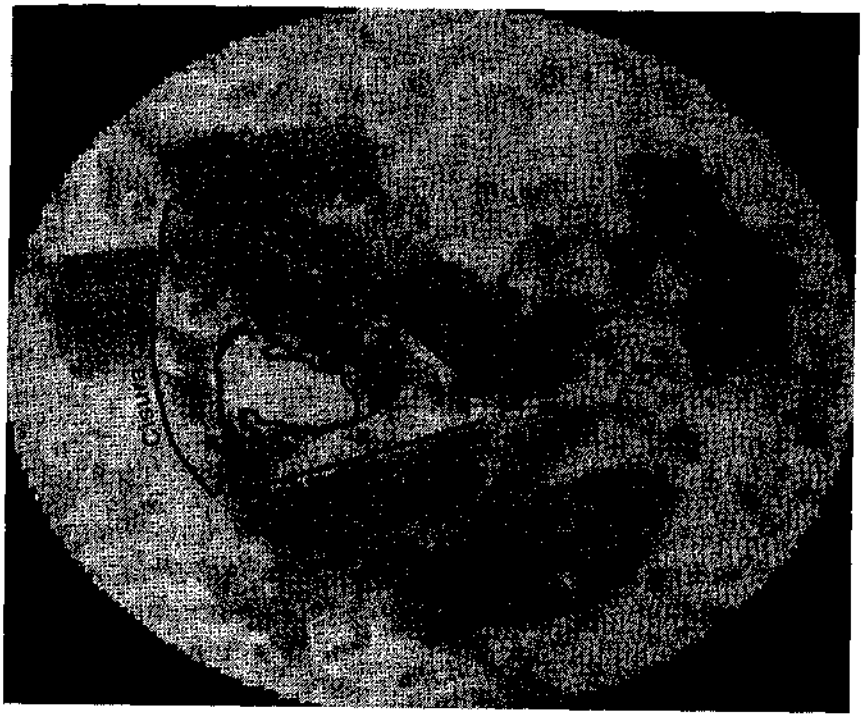

FIGURA 4: La imagen de contraste dependiente del nivel de oxigenación en sangre (BOLD), que probablemente refleja la actividad sináptica, medido mediante resonancia magnética funcional (RMf) en el interior de una zona de proyección de la lesión, no cambia en función del tiempo después de una lesión retinal. Se muestra un área de la corteza visual V_1 de un radio de 3 cm, centrada en las inmediaciones de la representación foveal. En esta sección aplanada se indica la zona delimitada por la cisura calcarina, el surco semilunar y el occipital inferior. Las áreas externas a ésta corresponden en gran medida a la corteza no visual. Los límites de la zona de proyección de la lesión se muestran a los 0 días (contorno interior), a los 4 meses (contorno exterior) y a los 7,5 meses (contorno intermedio) posteriores a la lesión. La zona de proyección de la lesión media, respectivamente, 158, 179 y 180 mm^2 (de la figura 2 en Smirnakis y otros, 2005).

mio Nobel, y Gilbert son neurocientíficos de primer orden. Pero tras quince años de estudios intensivos no se ha llegado a un consenso sobre la muy importante cuestión de si las redes neuronales corticales visuales primarias son plásticas (Giannikopoulos y Eysel, 2006). He expuesto con detalle esta historia no para repartir culpas, sino para insistir en que la complejidad biológica de los sistemas re-

ticulares sinápticos que intentamos comprender es muy considerable, y pone a prueba la destreza de incluso los mejores neurocientíficos. No obstante, si no entendemos algunas de las propiedades fundamentales de este primer enclave transmisor de la corteza a otras áreas corticales relacionadas con la función visual, como las del lóbulo temporal cuya función normal es necesaria para el reconocimiento de las caras, es muy improbable que lleguemos a comprender las redes sinápticas subordinadas de nuestra capacidad visual en zonas como el lóbulo temporal.

Las redes del cerebelo

La corteza cerebelosa posee un conjunto de tipos de neuronas y de conexiones sinápticas cuyas disposición y función parecen ser particularmente simples, como se muestra en la figura 5. El cerebelo parece ser por ello especialmente indicado para el análisis experimental y el desarrollo de teorías empíricamente verificables, un tema que me resulta especialmente fascinante (véase, por ejemplo, Gibson, Robinson y Bennett, 1991). Fue el genial David Marr (1969) quien introdujo la idea de que la adquisición de nuevas destrezas motrices depende de la plasticidad de las sinapsis entre los terminales axónicos paralelos y las células de Purkinje del cerebelo (fig. 5). Esta observación ha resultado extremadamente fructífera, pues en los treinta y cinco años siguientes la plasticidad de estas sinapsis se ha utilizado para explicar la adquisición y la retención durante toda la vida de muchas destrezas motoras, entre ellas la sincronización motora aprendida y la adaptación refleja (Ito, 2001). Se ha reunido un extenso cuerpo de estudios sobre el modelo de red sináptica en el que la actividad de los terminales axónicos de fibras trepadoras sobre las células de Purkinje re-

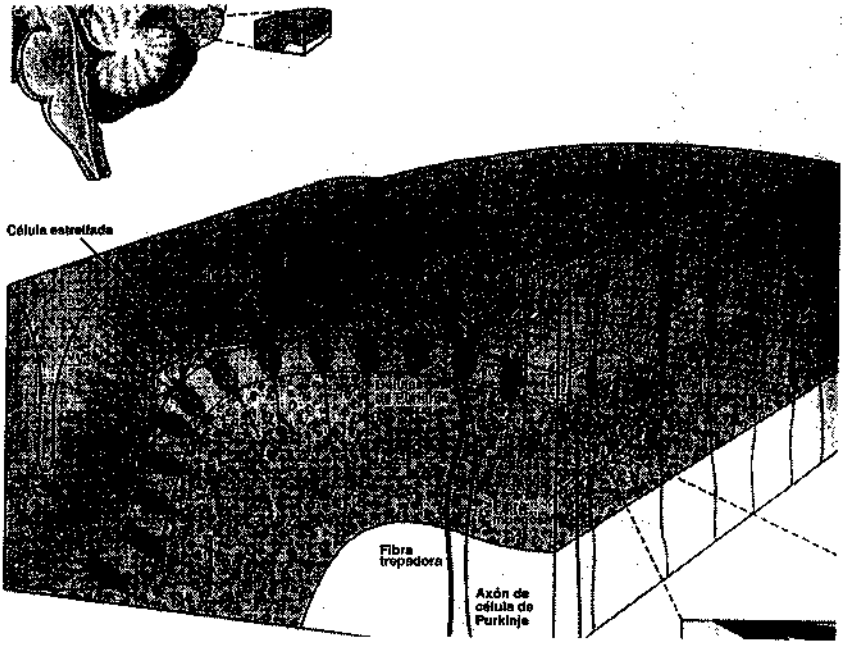

FIGURA 5: Córtex cerebeloso. Se muestran las grandes células de Purkinje con sus enormes árboles dendríticos. Cada uno de éstos recibe conexiones sinápticas de terminales axónicos de fibras paralelas que derivan de células granulares muy pequeñas que, a su vez, reciben sinapsis de terminales axónicos de fibras musgosas. Es de notar que, además, las células de Purkinje reciben conexiones sinápticas de un único axón de una fibra trepadora. Esta circuitería está dispuesta en un hermoso patrón repetitivo y regular, que hace del córtex cerebeloso un objeto ideal de investigación experimental. También se muestran neuronas inhibidoras (concretamente, células cesta, células estrelladas y células de Golgi).

duce la fuerza de las sinapsis de los terminales axónicos de fibras paralelas sobre estas mismas células cuando los dos inputs (es decir, los de las fibras trepadoras y los de las fibras paralelas) son estimulados a la vez (fig. 5). Se dice que esta forma de plasticidad sináptica está en la base de la adquisición de nuevas destrezas motoras. Si se utiliza la estimulación eléctrica para estimular directa y simultáneamente los terminales axónicos de fibras trepadoras y los de fibras paralelas, se reduce la amplitud de los potenciales sinápticos activados por los terminales axónicos de fibras paralelas sobre células de Purkinje (Ito y Kano,

1982). Esta depresión de los potenciales sinápticos requiere emparejamientos repetidos de inputs sinápticos de terminales axónicos de fibras trepadoras y de fibras paralelas. La depresión se mantiene durante muchas horas después de que concluya este protocolo de estimulación, y se denomina depresión a largo plazo. En los veintitrés años posteriores al descubrimiento de la depresión a largo plazo por parte de Ito y sus colegas, se han realizado una inmensa cantidad de investigaciones para averiguar su base molecular (Ito, 2002).

No obstante, un experimento llevado a cabo el año pasado causó una gran conmoción al demostrar que la depresión a largo plazo no interviene en el aprendizaje motor en el cerebelo. Llinas y sus colegas (Welsh y otros, 2005) impidieron por medios farmacológicos la depresión a largo plazo de las sinapsis formadas por terminales axónicos de fibras paralelas sobre células de Purkinje, después de la estimulación simultánea tanto de los terminales axónicos de fibras trepadoras como los de fibras paralelas (fig. 6). Esto no produjo efecto alguno en la adquisición de las destrezas motoras que intervienen en el test de la rueda giratoria (Lalonde, Bensoula y Filali, 1995), ni en el desarrollo de la sincronización motriz durante el condicionamiento del reflejo del parpadeo (McCormick y Thompson, 1984). Después de treinta y seis años de investigación, las redes sinápticas y los mecanismos moleculares implicados en el aprendizaje motor del cerebelo siguen sin esclarecerse.

La complejidad de la investigación sobre el sistema nervioso central

Los tres ejemplos anteriores muestran el progreso relativamente lento de la neurociencia en su intento de descifrar la complejidad incluso

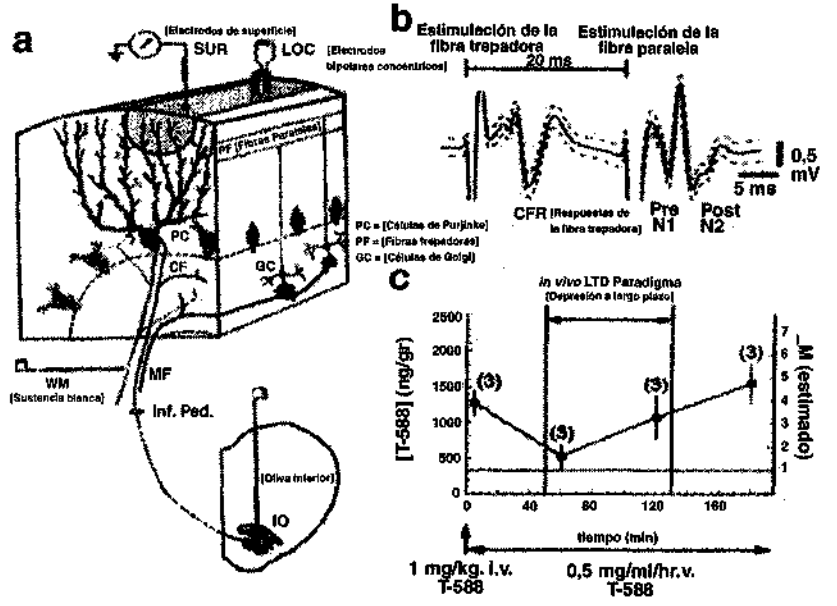

FIGURA 6: El fármaco T-588 impide la depresión a largo plazo de la fibra paralela a la sinapsis de la célula de Purkinje inducida por la estimulación *in vivo* simultánea de la fibra trepadora y la fibra paralela. Concentraciones similares de T-588 en el cerebro de ratones y ratas activos no afectaron ni al aprendizaje motor en el test de la rueda giratoria, ni al aprendizaje de la sincronización motora durante el condicionamiento clásico del reflejo del parpadeo. Así pues, la depresión a largo plazo de fibras paralelas a células de Purkinje bajo el control de fibras trepadoras no es necesaria para la adaptación motora ni para el aprendizaje de la sincronización de respuesta en dos modelos comunes de aprendizaje motor con implicación del cerebelo. En a) se muestran electrodos estimuladores bipolares concéntricos situados sobre la superficie cerebelosa (LOC) para estimular un haz de fibras paralelas (PFs) y dentro la sustancia blanca cerebelosa (WM) para estimular las fibras trepadoras (CFs). Se utilizó un electrodo de superficie de disco de plata (SUR) para registrar los potenciales de campo evocados en las células de Purkinje (PCs). Se empleó la estimulación directa de la oliva cerebelosa inferior (IO) para verificar que los potenciales generados por el electrodo WM eran respuestas de la fibra trepadora (CFR). b) muestra las CFR y las respuestas PF presinápticas (N_1) y postsinápticas (N_2) generadas por la estimulación conjunta de CF y PF utilizando un intervalo de 20 m entre los estímulos para generar una depresión a largo plazo de la respuesta N_2 PF. Las líneas de puntos indican una desviación estándar. En c) se muestran los resultados de las diferentes concentraciones de T-588 en el cerebro. Se destacan cuatro puntos temporales diferentes de la inyección intravenosa continua de T-588 con el paradigma de depresión a largo plazo (LTD), situados entre 50 y 130 m después del inicio de la inyección. La línea horizontal en 1 μM indica la concentración de T-588 que impedía la LTD *in vitro*. El número de cerebros de la muestra en cada punto temporal se indica entre paréntesis (extraído de la figura 1 de Welsh y otros, 2005).

de las partes «más simples» del sistema nervioso central. Sugieren que conviene desconfiar del endiosamiento tras la convicción de que los neurocientíficos comprenden muchas de las funciones del sistema nervioso central (si es que comprenden alguna en absoluto), y enfatizan hasta qué punto es necesario detenerse a reflexionar antes de aceptar muchas de las afirmaciones acerca de lo que hacen las redes sinápticas del cerebro.

Segunda fuente principal de interés relativa a nuestra comprensión de la función de las redes celulares de la corteza

He mencionado dos aspectos del enfoque ingenieril de la comprensión de las redes sinápticas que sugieren una pausa para la reflexión. He considerado el primero de ellos en el apartado segundo de este capítulo, concretamente la gran dificultad para determinar las propiedades biológicamente relevantes que resulten pertinentes para la construcción de un tipo ingenieril de red que pueda usarse para comprender mejor la red biológica. La segunda dificultad aparece cuando los atributos que normalmente adscribimos a los seres humanos, y en algunos casos a otros animales, se atribuyen a las redes sinápticas, sea antes o después de que hayan sido reducidos a dispositivos de ingeniería con diversos grados de complejidad y adaptabilidad. Se dice que determinadas redes sinápticas o grupos de redes sinápticas del cerebro recuerdan, ven y oyen. Por ejemplo, se señala que «podemos considerar toda visión como una búsqueda continua de las respuestas a preguntas que el cerebro plantea. Las señales procedentes de la retina constituyen "mensajes" que contienen esas respuestas» (Young, 1978). Se dice que la corteza visual del polo occipital (fig. 7)

posee neuronas que «proporcionan argumentos en base a los cuales el cerebro construye sus hipótesis de percepción» (Blakemore, 1977). Por lo que se refiere a las zonas del cerebro que deben funcionar para que podamos ver los colores, se afirma que tales zonas intervienen en «la interpretación que el cerebro hace de la propiedad física de los objetos (su reflectancia), una interpretación que le permite adquirir rápidamente conocimientos sobre la propiedad de la reflectancia» (Zeki, 1999).

No sólo se predican varios atributos psicológicos de haces de redes sinápticas del cerebro, sino también se predican de los hemisferios cerebrales enteros (fig. 7). Por ejemplo, se sugiere que «el hemisferio derecho es capaz de entender el lenguaje pero no la sintaxis» y que «la capacidad del hemisferio derecho para hacer inferencias es extremadamente limitada» (Gazzaniga, Ivry y Mangun, 2002). Y se dice además que «el hemisferio izquierdo también puede percibir y reconocer caras, y exhibir capacidades de reconocimiento superiores cuando se trata de caras familiares», y que «el hemisferio izquierdo adopta una útil estrategia cognitiva en la resolución de problemas, pero el hemisferio derecho carece de destrezas cognitivas adicionales» (Gazzaniga, Ivry y Mangun, 2002).

Estas afirmaciones de que las redes sinápticas, sean del tipo biológico o las resultantes de una útil reducción a un dispositivo ingenieril, poseen atributos psicológicos, me chocaron por extraordinarias. El lento y penoso progreso de la neurociencia en el uso del enfoque ingenieril para ilustrar las redes sinápticas no implicaba lógicamente que la teoría de que poseen atributos psicológicos no pueda mantenerse, pero me llevó a desconfiar de ella. Así que busqué ayuda para llegar a un esclarecimiento conceptual en aquellos estudiosos que están formados profesionalmente en tales materias, concretamente en filósofos. Después de leer obras de algunas de las figu-

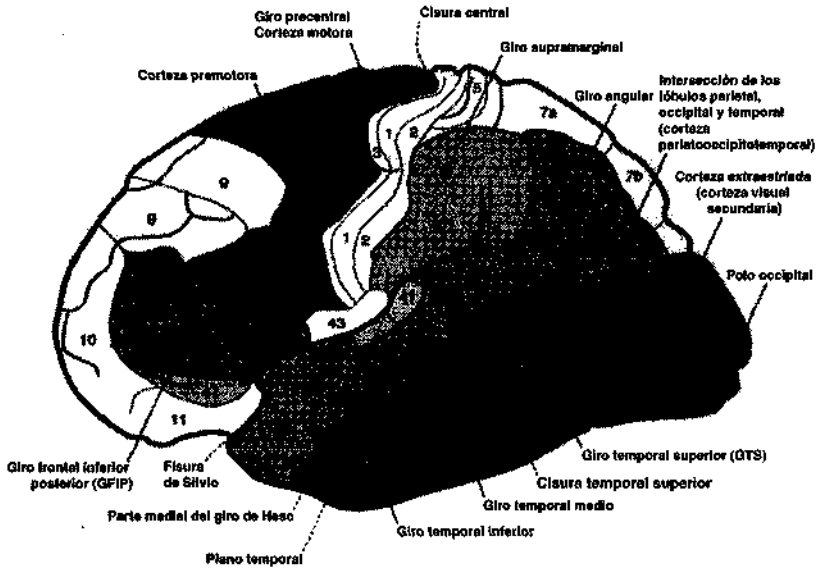

FIGURA 7: Corteza cerebral. Se muestra una visión lateral con las zonas numeradas que diseñó Brodmann (1909) como indicadoras de clases celulares distinguibles y las relaciones entre ellas. Las redes celulares referentes a la visión, por ejemplo, se encuentran en el polo occipital.

ras más importantes de la filosofía del siglo pasado, como Russell, Wittgenstein y Quine, inicié una discusión con algunos filósofos actuales, en particular con Peter Hacker, de Oxford. Nuestro diálogo sobre el tema de si los atributos psicológicos se podrían adscribir a las redes sinápticas se realizó exclusivamente a través de Internet y se completó antes de conocernos personalmente. Para mí ha sido un viaje inmensamente satisfactorio. Me ha obligado a reconsiderar la historia de la neurociencia desde Galeno, en el siglo II, hasta hoy, y a emprender junto con Peter un análisis crítico de las opiniones de los grandes de esta disciplina que han llevado a los neurocientíficos a las dificultades con que hoy se encuentran. Fruto de ese diálogo es nuestro libro *Philosophical Foundations of Neuroscience*. Mi satisfacción por este esfuerzo ha sido aún mayor por la invitación de la Ame-

rican Philosophical Association a que participe en un debate crítico con los profesores Dennett y Searle sobre las tesis de nuestro libro. El resultado ha sido un mayor esclarecimiento y, con ello, una mejora en nuestro intento por establecer la verdad en lo que se refiere a lo que las neurociencias han establecido y pueden confiar en alcanzar en el futuro. De este modo, hemos contribuido a favorecer los objetivos de la neurociencia de ayudar a comprender qué significa ser humano y paliar el sufrimiento de las personas.

Referencia bibliográfica

Barlow, H., «The neuron doctrine in perception», en M. S. Gazzaniga (comp.), *The New Cognitive Neurosciences*, 4ª ed., Cambridge, MIT Press, 1997, pág. 421.

Blake, R. y N. K. Logothetis, «Visual competition», en *Nature Reviews Neuroscience*, nº 3, 2002, págs. 13-21.

Crick, F. y C. Koch, «A framework for consciousness», en *Nature Reviews Neuroscience*, nº 6, 2003, págs. 119-126.

De Lima, A. D., T. Voigt y J. H. Morrison, «Morphology of the cells within the inferior temporal gyrus that project to the prefrontal cortex in the macaque monkey», en *Journal of Comparative Neurology*, nº 296, 1990, págs. 159-172.

Koch, C., *The Quest for Consciousness*, Greenwood Village, CO, Roberts, 2004.

Leopold, D. A. y N. K. Logothetis, «Multistable phenomena: changing views in perception», en *Trends in Cognitive Science*, nº 3, 1999, págs. 254-264.

Lumer, E. D., K. J. Friston y G. Rees, «Neural correlates of perceptual rivalry in the human brain», en *Science*, nº 280, 1998, págs. 1.930-1.934.

Srinivasan, R., D. P. Russell, G. M. Edelman y G. Tononi, «Increased synchronization of neuromagnetic responses during conscious perception», en *Journal of Neuroscience*, nº 19, 1999, págs. 5.435-5.448.

Tononi, G. y G. M. Edelman, «Consciousness and complexity», en *Science*, nº 282, 1998, págs. 1.846-1.851

LAS REFUTACIONES

LA FILOSOFÍA COMO ANTROPOLOGÍA INGENUA

Comentario sobre Bennett y Hacker

DANIEL DENNETT

El libro *Philosophical Foundations of Neuroscience*, de Bennett y Hacker (Blackwell, 2003), un trabajo de colaboración entre un filósofo (Hacker) y un neurocientífico (Bennett), es un ambicioso intento de reformular los objetivos de la investigación en neurociencia cognitiva mediante la demostración de que los científicos cognitivos y otros teóricos, incluido yo mismo, han estado confundiéndose mutuamente al usar el lenguaje de forma sistemáticamente «incoherente» y conceptualmente «confusa». Tanto en el estilo como en la sustancia, el libro evoca el Oxford de principios de la década de 1960, cuando la Filosofía del Lenguaje Corriente era hegemónica, y Ryle y Wittgenstein eran las autoridades en lo referente a los significados de nuestros términos mentalistas o psicológicos cotidianos. Yo mismo soy producto de esa época y ese lugar (como, en este sentido, lo es Searle) y, antes de pasar a mis críticas, que serán duras, quiero destacar lo que creo que es exactamente correcto de su enfoque: las lecciones a menudo olvidadas de la Filosofía del Lenguaje Corriente.

> Los estudios neurocientíficos [...] lindan con los psicológicos, y la claridad en lo que se refiere a los logros de la investigación del cerebro

> presupone claridad en lo que se refiere a las categorías de la descripción psicológica corriente: es decir, las categorías de sensación y percepción, cognición y recuerdo, cogitación e imaginación, emoción y volición. En la medida en que los neurocientíficos no distinguen los lindes de las categorías relevantes, corren el riesgo no sólo de formular las preguntas equivocadas, sino también de interpretar erróneamente sus propios resultados experimentales.
>
> (pág. 115)

Dicho queda.[1] Cuando los neurocientíficos hacen uso de los términos corrientes que constituyen el acervo que bauticé como «psicología popular»,[2] deben proceder con la mayor cautela, ya que estos términos tienen presuposiciones de uso que pueden subvertir sus propósitos y convertir lo que de otro modo serían teorías y modelos empíricos prometedores en sinsentidos apenas disimulados. El filósofo, experto en los matices de significado que cautivan la imaginación del teórico, es el tipo adecuado de pensador para llevar a cabo este importante ejercicio de higiene conceptual.

También convengo con ellos (aunque no lo expresaría en sus términos) en que «la evidencia en la que se basa la adscripción de atributos psicológicos a los demás no es de tipo inductivo, sino de criterio; la evidencia es evidencia lógicamente correcta» (pág. 82). Esto me coloca de su lado en contra, por ejemplo, de Fodor.[3]

Así pues, estoy de acuerdo sin reservas con el supuesto motivador de su libro. También comparto algunos de los blancos principales de su crítica, en particular su afirmación de que en la neurociencia cognitiva hay residuos cartesianos esparcidos por doquier, que provocan daños sustanciales. Dicen, por ejemplo:

> En general, los neurocientíficos contemporáneos consideran que los colores, los sonidos, los olores y los sabores son «construcciones mentales creadas en el cerebro mediante procesamiento sensorial. No existen, como tales, fuera del cerebro» [cita de Kandel y otros, 1995]. Esto, una vez más, difiere del cartesianismo sólo por la sustitución de la mente por el cerebro.
>
> (pág. 113)

Aquí critican un caso de lo que he llamado «materialismo cartesiano» (*Consciousness Explained*, 1991 [trad. cast.: *La conciencia explicada*, Barcelona Paidós, 1995]), y, en mi opinión, están en lo cierto al pensar que a muchos neurocientíficos cognitivos les deslumbra la idea de un lugar en el cerebro (que yo llamo el Teatro Cartesiano) donde se representa un espectáculo interior de notables construcciones ante una *res cogitans* (material) que constituye el público.

Más en concreto, creo que tienen razón al denunciar el cartesianismo atrofiante de la concepción de Benjamin Libet de acción intencional, y de parte de la obra teórica de Stephen Kosslyn sobre la imaginería mental. Coincido también con ellos en hallar deplorable el término «técnico» filosófico de los *qualia*, un regalo envenenado a la neurociencia donde los haya, y comparto algunos de sus recelos ante la expresión «cómo se siente» que Brian Farrell (1950) fue el primero en explorar, y que Thomas Nagel (1974) hizo famosa. La introspección, afirman, no es una forma de visión interior; no existe un ojo de la mente. Estoy de acuerdo. Y tener un dolor no es lo mismo que tener un penique; el dolor *no es una cosa* que esté ahí dentro. En efecto. Aunque no estoy de acuerdo con todo lo que dicen a lo largo de los derroteros por los que llegan a todos estos destinos, coincido con sus conclusiones. O, más exactamente, ellos coinciden con mis conclusiones, aunque no las citen.[4]

Más que su incapacidad de reconocer estos puntos realmente sustanciales de acuerdo, me sorprende que el núcleo de su libro, que es también el núcleo del ataque notablemente ofensivo que me dirigen,[5] sea un argumento que yo mismo concebí y del que me ocupé en profundidad ya en 1969. Ahí va lo que ellos llaman la falacia mereológica:

> Sabemos lo que supone para los seres humanos experimentar cosas, ver cosas, saber o creer cosas, tomar decisiones, interpretar datos equívocos, conjeturar y formular hipótesis. ¿Pero sabemos qué supone para un *cerebro* ver u oír, tener experiencias, saber o creer algo? ¿Tenemos idea de qué podría suponer para un *cerebro* tomar una decisión?
>
> Responden con un sonoro ¡NO!
>
> No tiene sentido atribuir predicados psicológicos (o sus negaciones) al cerebro, salvo de modo metafórico o metonímico. La combinación resultante de palabras no dice algo que sea falso; más bien no dice absolutamente nada, pues carece de sentido. Los predicados psicológicos son predicados que se aplican esencialmente al animal viviente en su totalidad, no a sus partes. No es el ojo (y mucho menos el cerebro) el que ve, sino que *nosotros* vemos *con* nuestros ojos (y no vemos con nuestro cerebro, aunque sin un cerebro con un sistema visual que funcione con normalidad, no veríamos).
>
> (pág. 72)

Esto es, cuando menos, muy cercano al argumento que expuse en 1969, cuando distinguía los niveles personal y subpersonal de la explicación. *Yo* siento dolor; no mi cerebro. *Yo* veo cosas; no mis ojos. Hablando del dolor, por ejemplo, señalaba:

> Un análisis de nuestra forma corriente de hablar sobre el dolor demuestra que no se pueden descubrir en el cerebro acontecimientos ni procesos que muestren las características de los supuestos «fenómenos mentales» del dolor, porque el modo de hablar de los dolores es esencialmente no mecánico, y los acontecimientos y los procesos del cerebro son esencialmente mecánicos.
>
> (*Content and Consciousness*, pág. 91)

Pese a lo mucho que tenemos en común, Bennett y Hacker menosprecian profundamente mi obra. ¿Cómo explicarlo? Como tan a menudo suele ocurrir en la filosofía, es una ayuda contar con alguien que diga, con rotundidad y claridad, lo que otros sólo insinúan o presuponen tácitamente. Bennett y Hacker se las arreglan para expresar posiciones que llevo combatiendo *indirectamente* cuarenta años, pero nunca había podido atacar de frente, por falta de un exponente manifiesto. Al igual que Jerry Fodor, en quien he confiado durante años para que expresara con contundencia *justamente* los puntos que yo deseaba negar —ahorrándome atacar a un hombre de paja—, Bennett y Hacker me proporcionan una doctrina clara que puedo criticar abiertamente. La tarea de poner en orden mis pensamientos sobre estos temas como reacción a sus afirmaciones me ha sido muy esclarecedora, como confío que lo sea también para otros.

El trasfondo filosófico

En este apartado voy a hablar sólo de Hacker, dejando a su coautor Bennett al margen de la discusión, ya que los puntos que voy a criticar son claramente aportación de Hacker. Repite, a menudo con las

mismas palabras, afirmaciones que hace en su libro *Wittgenstein: Meaning and Mind* (Blackwell, 1990), unas afirmaciones que son estrictamente filosóficas.

Cuando Hacker arremete contra mí, una y otra vez, por no conseguir apreciar la falacia mereológica, es como quien pretende enseñar a sorber huevos a su abuela. Estoy familiarizado con el argumento, pues fui el primero en invocarlo. ¿Es que, quizá, perdí el rumbo cuando me fui de Oxford? Entre los filósofos que han asumido mi distinción entre nivel personal y nivel subpersonal, una al menos (Jennifer Hornsby) ha conjeturado que tal vez abandoné ese camino en mi obra posterior.[6] ¿Es que realmente di la espalda a esa buena idea? No.[7] En esta ocasión sería de lo más adecuado citar mi crítica de 1980 a la defensa que Searle hace de la intuición tras la Habitación China:

> La respuesta de los sistemas indica, en mi opinión de forma completamente correcta, que Searle ha confundido los diferentes niveles de explicación (y atribución). *Yo* entiendo el inglés, no mi cerebro —ni, más en concreto, la parte correspondiente de mi cerebro (si es que se puede aislar) que «procesa» las frases percibidas y traduce mis intenciones comunicativas en actos de habla.
>
> (*Behavioral and Brain Sciences* [1980], 3, pág. 429)[8]

(Por cierto, Searle desechó sumariamente esta afirmación mía en su réplica en BBS. Tengo curiosidad por ver qué hace de la distinción entre nivel personal y nivel subpersonal bajo su forma de falacia mereológica.)[9]

El texto canónico en el que Hacker basa su convicción sobre la falacia mereológica es una frase suelta de San Ludwig:

> En resumen: Sólo de un ser humano y de lo que se parece a un ser humano (se comporta como tal) se puede decir: tiene sensaciones; ve, es ciego; oye, es sordo; es consciente o inconsciente.
>
> (*Philosophical Investigations*, párraf. 281)*

Aquí precisamente es donde Hacker y yo diferimos. Me complace citar estas palabras de Wittgenstein; de hecho considero que yo *extiendo* la posición de Wittgenstein: veo que los robots y los ordenadores que juegan al ajedrez y, sí, el cerebro y sus partes *realmente* «se parecen a un ser humano viviente (porque se comportan como un ser humano)», y esta semejanza es suficiente para garantizar un uso ajustado del vocabulario psicológico para caracterizar esa conducta. Hacker no se da cuenta de ello, y él y Bennett califican de «incoherentes» a todos los casos de tal uso, e insisten una y otra vez en que «no tienen sentido». Entonces, ¿quién tiene razón?

Regresemos a 1969 y veamos cómo yo exponía el tema entonces:

> En un sentido, la distinción entre los niveles personal y subpersonal no es nueva en absoluto. La filosofía de la mente que iniciaron Ryle y Wittgenstein es en gran medida un análisis de los conceptos que empleamos en el nivel personal, y la lección que hay que aprender de los ataques de Ryle a las «hipótesis paramecánicas» y a la insistencia a menudo sorprendente de Wittgenstein en que las explicaciones llegan a un fin bastante antes de lo que habíamos pensado, es que no hay que confundir los niveles personal y subpersonal. Sin embargo, en algunos casos la lección se ha comprendido mal, al igual que la lección de que **el nivel personal de la explicación es el único nivel de explicación cuando el objeto de estudio son las mentes y las acciones humanas**. En un sentido importante pero limitado, esto es verdad, pues,

* Trad. cast.: *Investigaciones filosóficas*, Barcelona, Altaza, 1999. *(N. del t.)*

como vemos en el caso del dolor, abandonar el nivel personal significa dejar de hablar del dolor. En otro sentido importante, es falso, y esto es lo que a menudo se olvida. El reconocimiento de que hay dos niveles de explicación da origen a la obligación de relacionarlos, y **ésta es una tarea que no se halla fuera de la provincia del filósofo** [...] Queda pendiente la cuestión de cómo cada porción del *discurso* sobre el dolor se relaciona con impulsos neurales o con el discurso sobre impulsos neurales. Esta pregunta y otras paralelas sobre otros fenómenos necesitan respuestas detalladas incluso después de haberse puesto de acuerdo en que hay diferentes tipos de explicación, diferentes niveles y categorías.

(*Content and Consciousness*, págs. 95-96)

Este pasaje da una idea general de la tarea que me he impuesto en los últimos treinta y cinco años. Y los fragmentos en negrita señalan los principales puntos de desacuerdo con Hacker, ya que mi camino no es en modo alguno el que él ha seguido. Él da sus razones, que merecen una detenida atención:

a) Las preguntas conceptuales anteceden a las cuestiones de verdad y falsedad [...] Se sigue que las preguntas conceptuales no se pueden trasladar a la investigación y la experimentación científicas ni al proceso de teorización científica.

(pág. 2)

Podemos cuestionar la primera afirmación. ¿Es que las respuestas a esas preguntas conceptuales no son o verdaderas o falsas? Según Hacker, no.

b) Lo que la verdad y la falsedad son a la ciencia, el sentido y el sinsentido son a la filosofía.

(pág. 6)

Así pues, cuando los filósofos cometen errores producen sinsentidos, nunca falsedades, y cuando hacen un buen trabajo no debemos decir que están en lo *cierto* ni que digan la *verdad*, sino simplemente que lo que dicen tiene sentido.[10] Me inclino a pensar que la proposición b) de Hacker es simplemente *falsa*, no un sinsentido, pero, sea lo que fuere, la segunda afirmación de Hacker en a), a pesar del «se sigue que», no se sigue. Aun en el caso de que las preguntas conceptuales «antecedan» a los temas de la verdad y la falsedad, es posible que todo aquel que desee dejar claro cuáles son las respuestas buenas deba investigar las indagaciones científicas relevantes de forma asidua. Hacker califica esta proposición de naturalismo quineano* y la rechaza con un argumento irrelevante: «no creemos que la investigación empírica pueda resolver cualquier problema filosófico, del mismo modo en que no puede resolver problemas de matemáticas» (pág. 414). Evidentemente, no; la investigación empírica no los *resuelve*, los *informa* y a veces los *ajusta* o *revisa*, y entonces esos problemas a veces se *disuelven*, y a veces se pueden solucionar mediante reflexión filosófica ulterior.

La insistencia de Hacker en que la filosofía es una disciplina *a priori* que no tiene continuidad con la ciencia empírica es la principal fuente de los problemas que aquejan este proyecto, como veremos:

> c) ¿Cómo se pueden investigar los límites del sentido? Sólo mediante el examen del uso de las palabras. El sinsentido a menudo aparece cuando una expresión se emplea en contra de las reglas que rigen su uso. La expresión en cuestión puede ser una expresión corriente, no técnica, en cuyo caso las reglas para su uso se pueden deducir de su

* En referencia al filósofo y matemático norteamericano Willard V. O. Quine (1908-2000), que cuestionó la existencia de verdades conceptuales no empíricas. *(N. del t.)*

> empleo habitual y de las explicaciones estándar de su significado. O puede ser un término técnico de una disciplina, en cuyo caso las reglas para su uso deben deducirse de la introducción que el teórico haga del término y de las explicaciones que ofrezca del uso estipulado. Ambos tipos de términos se pueden usar indebidamente, y cuando eso ocurre se produce el *sinsentido*, una combinación de palabras que está excluida del lenguaje. Y es que o bien no se ha estipulado nada sobre lo que el término significa en el aberrante contexto en cuestión, o bien esta forma de las palabras está de hecho excluida por una regla que especifica que no existe nada así (por ejemplo, que no existe nada que sea el «este del Polo Norte»), que ésta es una combinación de palabras que no tiene uso.
>
> (pág. 6)

Este pasaje evoca en gran medida la década de 1960, y quiero recordar al lector algunos de los problemas que encierra, problemas que yo creía resueltos desde hace muchos años, si bien entonces no disponíamos de una versión tan manifiesta para usarla como blanco de nuestra crítica.

> ¿Cómo se pueden investigar los límites del sentido? Sólo mediante el examen del uso de las palabras.

Observemos, en primer lugar, que, con independencia de lo que cualquier filósofo pueda decir, el examen del uso de las palabras es una investigación empírica, que a menudo produce verdades y falsedades comunes y corrientes, y está sometida a la corrección por parte de observaciones y objeciones estándar. Tal vez fue una vaga apreciación de esta flagrante contradicción lo que llevó a Hacker a decir lo siguiente en su libro de 1990:

> La gramática es autónoma, no tiene que rendir cuentas a las proposiciones factuales, pero éstas deben presuponerla. En este sentido, a diferencia de las reglas de la lógica medios/fines, es arbitraria. Pero está emparentada con lo no arbitrario. Está moldeada por la naturaleza humana y la naturaleza del mundo que nos rodea.
>
> (pág. 148)

Dejemos que la gramática sea autónoma, signifique esto lo que signifique. Pese a ello, no se puede estudiar sin formular preguntas, e incluso en el caso de que uno se haga las preguntas *a sí mismo*, debe considerarse lo que se dice. La convicción de que este método de consultar las propias intuiciones (gramaticales o de otro tipo) es enteramente distinto de la indagación empírica tiene una larga tradición (que se remonta no sólo al Oxford de la década de 1960, sino a Sócrates), pero no resiste el análisis reflexivo.

Esto se puede ver perfectamente si comparamos este estilo de filosofía con la antropología, una indagación manifiestamente empírica que se puede hacer bien o mal. Si uno escoge informantes de segunda clase, o no domina primero su lengua, realizará un trabajo de tercera clase. Por esta razón, algunos antropólogos prefieren practicar una forma u otra de *autoantropología*, en la que uno se utiliza a sí mismo como informante, tal vez con la complicidad de algunos colegas cercanos como interlocutores. La naturaleza empírica de la empresa es exactamente la misma.[11] Los lingüistas, como es notorio, practican una forma de antropología de este tipo, y saben mucho, a estas alturas, de los escollos y los riesgos de sus particulares intentos de descifrar las intuiciones gramaticales relativas a sus lenguas maternas. Es bien sabido, por ejemplo, lo difícil que resulta no contaminar las propias intuiciones sobre gramaticalidad con ideas teóricas de nuestra predilección. De hecho, algunos lingüistas han llegado

a la conclusión de que los lingüistas teóricos no son, o no deberían ser, reclutados como informantes, ya que sus juicios no son inocentes. A Hacker y otros filósofos de ideas similares se les plantea aquí todo un reto: ¿cómo distinguen, exactamente, sus indagaciones de la autoantropología, una investigación empírica que manifiestamente usa los mismos métodos y lleva a juicios del mismo tipo?[12]

Cualquiera que piense que los filósofos han encontrado un método de investigación *gramatical* que de un modo u otro es inmune (o independiente, o «previa») a los problemas que se le pueden plantear a esa indagación antropológica, nos debe una explicación detallada de cómo lo hace. Simples afirmaciones de que *esto es lo que hacen los filósofos* no hacen más que evadir la cuestión. Mi colega Avner Baz me recuerda una interesante sugerencia de Stanley Cavell[13] para afrontar esta dificultad: Cavell sostiene que las observaciones del filósofo sobre lo que decimos son más próximas a los juicios *estéticos*. En palabras de Baz: «uno presenta su juicio como *ejemplar*, habla *para* una comunidad» (comunicación personal), y esto funciona hasta cierto punto, pero dado que el antropólogo también trata de hallar la interpretación *mejor* y *más coherente* de los datos disponibles (la idea de Quine sobre el principio de caridad, y la mía sobre el supuesto de racionalidad de la actitud intencional), este elemento normativo o sancionador está ya presente (aunque entre paréntesis) en la investigación del antropólogo. Éste no puede interpretar lo que sus informantes *dicen* sin desvelar lo que *deberían decir* (en su comunidad) en muchas situaciones. Sin embargo, lo que deliberadamente se deja al margen de la empresa del antropólogo, y es necesario defender en la del filósofo, es una *justificación* de la afirmación siguiente: Esto es lo que esta gente hace y dice, *y todos deberíamos hacer lo mismo*. Como veremos, Hacker no consigue identificar la comunidad por la que habla y ello malogra su proyecto.

Volvamos a c):

> El sinsentido a menudo aparece cuando una expresión se emplea en contra de las reglas que rigen su uso.

Hace demasiado tiempo que este tipo de subterfugio filosófico debería haberse proscrito. Como es notorio, Ryle afirmaba haber identificado «errores categoriales» apelando para ello a la «lógica» de las afirmaciones de existencia, pero afrontémoslo: se trataba de un farol. Carecía de una lógica articulada de los términos de existencia que fundamentara sus afirmaciones. Pese a la popularidad de este tipo de discurso, de Ryle y Wittgenstein a toda una hueste de imitadores, ningún filósofo ha formulado *nunca* «las reglas» para el uso de ninguna expresión corriente. Los filósofos han emitido cientos de juicios sobre usos desviados, pero señalar que «nosotros no lo diríamos de esta guisa» no equivale a expresar una regla. Los lingüistas emplean un asterisco o una estrella para hacer el mismo tipo de observación, y han engendrado miles de oraciones con asterisco como las siguientes:

* Una bellota creció como cada roble.
* La casa infestaba de ratas.

Pero, como cualquier lingüista nos diría, llamar la atención hacia un juicio de uso desviado —incluso cuando remite a un patrón de desviación importante y bien descrito— no es lo mismo que desvelar las reglas que rigen esos casos. Los lingüistas llevan más de cuarenta años trabajando con ahínco para formular las reglas de la sintaxis y la semántica de la lengua, y hay pequeñas áreas en las que pueden decir de forma convincente que han descubierto «las reglas».

Pero también se han encontrado con amplias zonas de vaguedad. ¿Qué diríamos de esta frase?

* El gato trepó árbol abajo [ejemplo de Jackendoff]

¿Se trata de un sinsentido que contraviene «las reglas» del verbo *trepar*? Es difícil decirlo, y pudiera ser que el uso esté cambiando. Los ejemplos de este tipo son abundantes. Los lingüistas han aprendido que algo puede sonar un tanto raro, dar mala espina, sin por ello violar ninguna regla clara que nadie haya podido enunciar y defender. Y la idea de unas *reglas* inefables es demasiado oscurantista para que merezca la pena discutirla. Las intuiciones de los filósofos, por bien afinadas que estén, no son una fuente superior de evidencia en esta investigación manifiestamente empírica.

Volvamos a c). Hacker pasa a dividir el léxico en dos:

> La expresión en cuestión puede ser una expresión corriente, no técnica, en cuyo caso las reglas para su uso se pueden deducir de su empleo habitual y de las explicaciones estándar de su significado. O puede ser un término técnico de una disciplina, en cuyo caso las reglas para su uso deben deducirse de la introducción que el teórico haga del término y de las explicaciones que ofrezca del uso estipulado.

Me siento tentado a afirmar que Hacker está sencillamente equivocado (pero no diciendo un contrasentido) cuando sugiere que el signo distintivo de un término técnico es que lo «introduce» un teórico que «estipula» su uso. O bien dice esto, o está definiendo «término técnico» de forma tan restringida que muchos términos que normalmente consideraríamos técnicos no lo serían para Hacker —y «término técnico» es un término técnico cuyo uso Hacker está

estipulando aquí y ahora—. Dejémosle, pues, con su definición de los términos técnicos, por restringida que sea. Ninguno de los términos que son objeto de sus ataques en el libro es técnico en este sentido, por consiguiente deben ser términos «corrientes, no técnicos», o bien son híbridos, una posibilidad que Hacker contempla brevemente y luego rechaza en su libro de 1990:

> Si los neurofisiólogos, los psicólogos, los expertos en inteligencia artificial o los filósofos desean cambiar la gramática existente para introducir nuevas formas de hablar, pueden hacerlo; pero deben explicar sus nuevas estipulaciones y definir las condiciones de aplicación. Lo que no se puede hacer es argumentar que, ya que sabemos lo que significan «pensar», «ver» o «inferir», y sabemos lo que significa «cerebro», entonces debemos saber qué significa que «el cerebro piensa, ve e infiere».Y es que sabemos lo que significan estos verbos sólo en la medida en que dominamos su uso existente, que no autoriza a aplicarlos al cuerpo o a sus partes, a no ser de modo derivativo. Tampoco se pueden mezclar el nuevo uso «técnico» y el antiguo, como por ejemplo suelen hacer los neurocientíficos cuando teorizan. La razón es que esto origina un conflicto de reglas y por consiguiente una incoherencia en el *uso que los neurocientíficos* hacen de estos términos.
>
> (págs. 148-149)

Esta última afirmación, que también está en el núcleo del libro de 2003, es una petición de principio. Si Hacker fuera capaz de *mostrarnos* las reglas, y de decirnos exactamente cómo los nuevos usos entran en conflicto con ellas, podríamos estar en condiciones de estar o no de acuerdo con él, pero sólo está especulando. No tiene ni idea de cuáles son «las reglas» para el uso de estos términos psicológicos cotidianos. Más revelador aún es que su obstinada insistencia en una

metodología apriorística le impide ver lo que está haciendo en este punto. Aceptemos que su convicción de que posee un método *a priori* que le da una intuición «previa» de los significados de sus términos psicológicos corrientes es *correcta*.[14] Sigue estando obligado a mostrar de qué forma su prolegómeno o preparación de la escena salvan el escollo de lo que podríamos llamar miopía conceptual: tratar los *propios* conceptos (posiblemente restingidos y mal fundamentados) como vinculantes para otros con diferentes objetivos y distinta formación. ¿Cómo establece que él y aquellos cuya obra critica hablan la misma lengua? Se trata sin duda de una cuestión empírica, y la incapacidad de Hacker de abordarla con el cuidado suficiente le ha desviado del camino. Lo que hace, en realidad, no es buena filosofía, sino mala antropología. Se dirigió a la ciencia cognitiva para «examinar el uso de las palabras» y no se dio cuenta de que él mismo llevaba *su* propio lenguaje corriente a un territorio ajeno, ni de que *sus* intuiciones no eran necesariamente aplicables. Cuando califica el uso de *ellos* de «aberrante», comete un error de principiante.

El uso de predicados psicológicos en la teorización de los científicos cognitivos es sin duda una jerga del inglés, muy distinta de la forma de hablar de los profesores de Oxford, y tiene sus propias «reglas». ¿Cómo lo sé? Porque he hecho el estudio antropológico. (Hay que ser un naturalista quineano para no cometer estos errores tan simples.) Hay un pasaje revelador en que Hacker reconoce esta posibilidad, pero muestra su incapacidad para tomársela en serio:

> ¿Es un descubrimiento nuevo que los cerebros también llevan a cabo tales actividades humanas? ¿O es una innovación lingüística que han introducido los neurocientíficos, los psicólogos y los científicos cognitivos, ampliando así el uso corriente de estas expresiones psico-

lógicas por buenas razones teóricas? ¿O, lo que sería más alarmante, se trata de una confusión conceptual?

(págs. 70-71)

Hacker opta por la tercera posibilidad, sin argumentarlo, mientras que yo me inclino por las dos primeras conjuntamente. *Hay* un elemento de descubrimiento. Es un hecho empírico, y sorprendente, que nuestro cerebro —más en concreto, *partes* de nuestro cerebro— lleva a cabo procesos *asombrosamente* parecidos a conjeturar, decidir, creer, llegar a conclusiones, etc. Y se parecen *lo suficiente* a estas conductas en el nivel personal para justificar la extensión del lenguaje corriente para abarcarlos. Si no se ha estudiado la excelente obra científica que esta adopción de la actitud intencional ha hecho posible, se pensará que es una locura hablar así. Pero no lo es.

De hecho, esto es lo que me llevó a desarrollar mi concepción de la actitud intencional. Cuando empecé a dedicar mi tiempo a hablar con investigadores en informática y neurociencia cognitiva, lo que me impresionó fue que, sin afectación alguna, sin dar golpecitos con el codo ni enarcar las cejas, hablaban de cómo los ordenadores (y programas y subrutinas y partes del cerebro, etc.) *quieren, piensan, concluyen, deciden*, etc. ¿Cuáles son las reglas?, me pregunté. Y la respuesta a la que llegué son las reglas para la adopción de la actitud intencional. La pregunta *factual* es: ¿hablan así las personas que trabajan en estos campos, y la actitud intencional capta al menos una parte esencial de «las reglas» que rigen esa forma de hablar? Y la respuesta (factual) es: *Sí*.[15] Hay también, supongo, una cuestión política: ¿tienen derecho a hablar así? Bueno, que lo hagan compensa con creces, porque ello les permite generar hipótesis verificables, desarrollar teorías, analizar fenómenos enormemente complejos en sus partes más comprensibles, etc.

Hacker también descubre este uso omnipresente de términos intencionales en la neurociencia, y está atónito, creedme, ¡atónito! Tantas personas cometiendo tamaños errores conceptuales. Y no sabe de la misa la mitad. No son sólo los neurocientíficos, son los informáticos (y no sólo los que trabajan en Inteligencia Artificial), los etólogos cognitivos, los biólogos celulares, los estudiosos de la evolución... todos aceptando el juego alegremente, enseñando a sus alumnos a pensar y hablar de esta forma, una pandemia lingüística. Si al ingeniero eléctrico medio le preguntáramos cómo funciona la mitad de los aparatos de nuestra casa, la respuesta estaría repleta de los términos intencionales que constituyen la falacia mereológica (si es que es una falacia).

No es una falacia. No atribuimos a las partes del cerebro una creencia (o una decisión o un deseo —o un dolor, sabe Dios—) *cabal*; esto sí *sería* una falacia. No; atribuimos a estas partes una forma atenuada de creencia y deseo, una creencia y un deseo despojados de muchas de sus connotaciones cotidianas (sobre la responsabilidad y la comprensión, por ejemplo). Del mismo modo que un niño pequeño puede *algo así como* creer que su papá es médico (sin tener una comprensión completa de lo que son un papá o un médico),[16] un robot —o cierta parte del cerebro de una persona— puede *algo así como* creer que hay una puerta abierta unos metros más adelante, o que algo pasa allá, por la derecha, etc. Hace años que defiendo tales usos de la actitud intencional para caracterizar sistemas complejos que van desde los ordenadores que juegan al ajedrez a los termostatos, y para caracterizar los subsistemas del cerebro en muchos niveles. La idea es que, cuando diseñamos un sistema complejo (o desciframos el diseño de un sistema biológico como una persona o el cerebro de una persona), podemos hacer progresos desmenuzando a la maravillosa persona como totalidad en ciertos tipos de subpersonas, sistemas agentivos

que tienen *parte* de la capacidad de una persona, y luego estos homúnculos se pueden desmenuzar en agentes aún más simples, aún menos persona, y así sucesivamente: una regresión *finita*, no infinita, que toca fondo cuando llegamos a agentes tan estúpidos que se pueden sustituir por una máquina. Es posible que todos mis intentos de justificar y explicar este movimiento se entiendan mal, pero dado que Bennett y Hacker nunca los consideran, no están en condiciones de valorarlos.

Así lo digo en un artículo que Hacker cita varias veces (aunque no este pasaje):

> Uno puede tener la tentación de preguntarse: ¿los componentes subpersonales son sistemas intencionales *auténticos*? ¿En qué punto de la disminución de la capacidad, a medida que descendemos hacia las simples neuronas, desaparece la intencionalidad *auténtica*? No hay que preguntárselo. Las razones para considerar que una neurona individual (o un termostato) es un sistema intencional son insignificantes, pero no cero, y la seguridad de nuestras atribuciones intencionales en los niveles más altos no depende de que identifiquemos un nivel inferior de intencionalidad auténtica.
>
> («Self-portrait», en Guttenplan, catalogado por H y B como «Dennett, Daniel C. Dennett», 1994)

La *falacia* del homúnculo, al atribuir *toda* la mente a una parte adecuada del sistema, no hace más que posponer el análisis y, por consiguiente, generaría un retroceso infinito, ya que cada postulación no produciría progreso alguno. Lejos de ser un *error*, atribuir una semi-demi-proto-cuasi-seudo intencionalidad a las partes mereológicas de las personas es precisamente el movimiento habilitador que nos permite ver cómo se obtienen maravillosas personas completas

a partir de estúpidas partes mecánicas. Se trata de algo endemoniadamente difícil de imaginar, y la licencia poética que permite la actitud intencional facilita sustancialmente la tarea.[17] Así pues, desde mi posición, Hacker es cómicamente ingenuo, como el gramático chapado a la antigua que reprende a quien utiliza la forma coloquial y dialectal inglesa *ain't* e insiste en que *no se dice así* a personas que saben lo que quieren decir cuando lo dicen. Hacker había previsto esta perspectiva en su libro de 1990 y la describió realmente muy bien:

> Si hay que tomar en serio todo esto [la ciencia cognitiva], parece que demuestra, en primer lugar, que la observación gramatical de que estos predicados, en su uso literal, están limitados a los seres humanos y a lo que se comporta como los seres humanos, o bien es simplemente errónea o bien exhibe una «inercia semántica» que ha sido superada por el avance de la ciencia, ya que las máquinas realmente se comportan como seres humanos. En segundo lugar, si tiene sentido literal atribuir predicados epistémicos e incluso perceptivos a máquinas diseñadas para simular ciertas funciones humanas y realizar determinadas tareas humanas, parece verosímil suponer que el cerebro humano debe tener una estructura funcional abstracta similar a la del diseño de la máquina. En este caso, no hay duda de que atribuir la diversidad de los predicados psicológicos al cerebro humano debe tener sentido después de todo.
>
> (págs. 160-161)

Exactamente. Ésa es la tesis. ¿Cómo la refuta? No lo hace. Dice: «Los problemas filosóficos derivan de la confusión conceptual. No los resuelven los descubrimientos empíricos, y el cambio conceptual no puede responder, sino sólo barrerlos y esconderlos debajo de la alfombra» (pág. 161). Dado que los problemas filosóficos de Hacker se están quedando obsoletos, supongo que podríamos limitarnos a

esconderlos debajo de la alfombra, pero preferiría darles una adecuada sepultura.

Los detalles neurocientíficos

Cuando Bennett y Hacker emprenden su examen de la literatura neurocientífica, sus críticas adolecen de poca variedad. Citan a Crick, Edelman, Damasio, Gregory y otros muchos diciendo cosas sorprendentemente «incoherentes», porque estos científicos caen en la llamada falacia mereológica:

> Las expresiones psicológicas que emplean, lejos de ser nuevos usos homónimos, se invocan en su sentido acostumbrado, de lo contrario los neurocientíficos no inferirían de ellas lo que infieren. Cuando Crick afirma que «lo que vemos no es *realmente* lo que hay; es lo que nuestro cerebro *cree* que hay», es importante que asuma que «cree» tiene sus connotaciones normales —que no significa lo mismo que cierto término nuevo «cree*»—. Porque forma parte de la teoría de Crick que la creencia es el resultado de una *interpretación* basada en la *experiencia* o la *información* previas (y no el resultado de una interpretación* basada en la experiencia* y la información* previas).
>
> (pág. 75)

De hecho Bennett y Hacker están simplemente equivocados (pero no incurren en un sinsentido). Crick pretende que toda su explicación (en este caso una explicación bastante trivial y en modo alguno polémica) se entienda en el nivel subpersonal. La interpretación en cuestión *no* es de la *experiencia* (en el nivel personal), sino, por ejemplo, de *los datos de la corriente ventral*, y se supone que el proceso

de interpretación es obviamente un proceso subpersonal. Otro pasaje con el mismo espíritu:

> Asimismo, cuando [J. Z.] Young afirma que el cerebro contiene conocimiento e información, que están codificados en el cerebro «del mismo modo que el conocimiento está registrado en libros u ordenadores», se refiere al conocimiento, no al conocimiento* —ya que son el conocimiento y la información, no el conocimiento* y la información*, lo que se puede registrar en libros y ordenadores.

Los autores no han hecho nada en absoluto para establecer que no existe un concepto de conocimiento o información que se pueda codificar tanto en los libros como en el cerebro. En la ciencia cognitiva, hay una literatura vasta y compleja sobre el concepto de información —y sobre el de conocimiento (basta pensar en las discusiones de Chomsky sobre el «cognoscimiento»* como respuesta a una crítica que iba muy en la misma línea que la de los autores)—, y el caso omiso que los autores hacen de estos debates anteriores demuestra que no se toman su tarea muy en serio. Se podrían citar muchos otros ejemplos del mismo estilo. Nuestros autores tienen una idea, la falacia mereológica, y la utilizan de forma sistemática, sin considerar los detalles en lo más mínimo. En cada ocasión citan el pasaje ilegítimo —y podrían haber encontrado muchísimos más ejemplos de atribuciones de actitud intencional a los subsistemas del cerebro—, y luego no hacen más que declararlo un contrasentido

* *Cognizing*: neologismo introducido por Chomsky para describir la relación entre un hablante y las reglas de su lenguaje y, en general, entre una persona y su conocimiento, y que denota un tipo de conocimiento en gran parte implícito o tácito. Véase Chomsky, N., *Knowledge of Language*, Nueva York, Praeger, 1986 (trad. cast.: *El conocimiento del lenguaje*, Madrid, Alianza, 1989). *(N. del t.)*

porque incurre en su falacia. Ni una sola vez intentan demostrar cómo el autor en cuestión, al cometer tan terrible error, se ve abocado irremediablemente a algún error o contradicción reales. ¿Quién podía imaginar que la filosofía de la neurociencia fuera tan fácil?

Consideremos su discusión del fascinante y polémico tema de la imaginería mental. En primer lugar, demuestran —aunque dudo que alguien lo haya puesto en cuestión alguna vez— que la imaginación creativa y la imaginería mental son en realidad fenómenos bastante distintos e independientes. Luego viene el tiro de gracia: «Una zona sensorial dispuesta topográficamente no es una imagen de nada; en el cerebro no hay imágenes, y el cerebro no *tiene* imágenes» (pág. 183). Como alguien que ha defendido con empeño durante años que no debemos precipitarnos en concluir que la «imaginería mental» implica imágenes reales en el cerebro, y que los haces retinotópicos que allí se encuentran es posible que no *funcionen* como imágenes en el procesamiento que realiza el cerebro, debo señalar que la pobre declaración de Bennett y Hacker no ayuda nada en absoluto. Es simplemente irrelevante si «dijéramoss» que el cerebro *tiene imágenes*. La pregunta de si algunos de los haces de estímulos del cerebro que manifiestamente tienen las propiedades geométricas de las imágenes realmente *funcionan* como imágenes es una pregunta empírica, y una pregunta cuya respuesta está próxima. El análisis filosófico no puede zanjar la cuestión —excepto mediante una insistencia profundamente reaccionaria en que estas estructuras icónicas de datos, de las que al parecer se extrae la información de forma muy similar a como las personas (en el nivel personal) extraemos información visualmente de las imágenes públicas, no *cuentan* como imágenes—. Tales arranques de ofuscación le generan a la filosofía un grave problema de credibilidad ante la ciencia cognitiva.

De hecho, existen graves problemas conceptuales en la forma en que los científicos cognitivos hablan de imágenes, conocimiento, representaciones, información y todo lo demás. Pero es una tarea difícil y pormenorizada la de *demostrar* que la terminología usada se emplea en modos que confunden gravemente a los teóricos. El hecho es que, en su mayor parte, estos términos, tal como se encuentran en la ciencia cognitiva, realmente son «lenguaje corriente» —no términos técnicos[18] que se hayan estipulado explícitamente en el seno de alguna teoría—. A menudo, a los teóricos les ha sido útil hablar, de forma *un tanto* impresionista, sobre información que se procesa, decisiones a las que se llega, representaciones que se consultan. Y, cuando se les cuestiona lo que dicen, en vez de hacer lo que haría un filósofo, es decir, *definir sus términos con mayor exactitud*, apuntan a sus modelos y dicen: «Miren: he aquí los mecanismos en acción, realizando el procesamiento de la información de que les hablaba». Y los modelos funcionan. Se *comportan* como deben comportarse para hacer aparecer aquel particular homúnculo, de modo que no haya ya necesidad de más reparos sobre qué se atribuía exactamente al sistema.

Pero también hay muchas ocasiones en que el entusiasmo de los teóricos por las interpretaciones intencionales de sus modelos los confunde.[19] Por ejemplo, en el debate sobre la imaginería, se han dado pasos en falso en forma de interpretaciones exageradas —por ejemplo por parte de Stephen Kosslyn— que es preciso rectificar. No es que hablar de mapas o imágenes en la neurociencia sea algo *completamente* fuera de lugar, sino que debe hacerse con mucho cuidado, algo que no siempre ocurre. ¿Puede ayudar la filosofía? Sí, dicen Bennett y Hacker: «Puede explicar —como nosotros hemos explicado— por qué las imágenes mentales no son como fotografías etéreas y por qué no se las puede rotar en el espacio mental» (pág. 405). Este enfoque maximalista no resulta útil. No se puede establecer qué

ocurre realmente en el cerebro cuando las personas están inmersas en la imaginería mental diciendo que el nivel personal no es el nivel subpersonal. Los teóricos ya lo saben; no cometen *ese* error. En realidad son pensadores muy cuidadosos y sutiles, y algunos de ellos *aún* quieren hablar de imágenes en el cerebro que funcionan como imágenes. Bien puede ser que estén en lo cierto.[20] Los filósofos como Hacker pueden perder interés en el tema una vez que se sitúa en el nivel subpersonal,[21] pero entonces no deberían cometer el error de criticar un ámbito del que saben tan poco.

A veces los autores se confunden con una seguridad tan cándida que el efecto resulta cómico, como en su severa reprobación de David Marr:

> Ver no significa descubrir algo a partir de una imagen o haz de luz que incide en la retina. Porque no es posible descubrir nada, en este sentido, a partir de algo que uno no puede percibir (no percibimos el haz de luz que incide en nuestras retinas, lo que percibimos es lo que sea que el haz de luz nos permite percibir).
>
> (pág. 144)

Entendido. Dado que *nosotros* no percibimos el haz de luz que da en nuestras retinas, es obvio que *nosotros* no descubrimos nada. Marr no era idiota. Entendía esto. Entonces, ¿qué pasa con la teoría de Marr sobre los *procesos subpersonales* de la visión?

> Además, *no está claro en absoluto* [la cursiva es mía] cómo el hecho de que la mente tenga acceso a las supuestas *descripciones* neurales capacitará a la persona para *ver*. Y si Marr insistiera (con toda la razón) en que es la persona, no la mente, la que ve, ¿cómo se explica la transición desde la presencia de una descripción de un modelo en tres dimensiones codificado en el cerebro a la experiencia de ver lo que

está ante los ojos? En realidad, *éste* no es un problema empírico que posteriores investigaciones deban resolver. Es el producto de una confusión conceptual, y lo que precisa es dilucidación.

(pág. 147)

Yo diría, al contrario, que es un problema filosófico que se resuelve haciendo entender a quienes piensan que «*no está claro en absoluto*» que la teoría de Marr *puede* explicar el repertorio de competencias asociadas a la visión.[22] Marr daba más o menos por supuesto que sus lectores podrían entender cómo un modelo del cerebro que dispone de una reproducción del mundo en tres dimensiones puede explicar cómo una criatura con precisamente este tipo de cerebro puede ver, pero si algunos lectores no conseguían entenderlo, un filósofo sería probablemente el especialista adecuado para explicárselo. Limitarse a *afirmar* que Marr padece de una confusión conceptual tiene, como tan bien lo expresó Russell, todas las ventajas del robo frente a un trabajo honrado.

Porque ver algo es el ejercicio de una capacidad, un uso de la facultad visual —*no* [la cursiva es mía] el procesamiento de información en sentido semántico ni la producción de una descripción en el cerebro.

(pág. 147)

Éste *no* es otro robo. Lo que se tiene que explicar es la capacidad de la «facultad visual», y esa capacidad se explica en términos de las capacidades combinadas de sus partes, entre cuyas actividades están la creación y la consulta de descripciones (si se las puede llamar así). Estos ejemplos se podrían multiplicar hasta el aburrimiento:

> No tiene sentido, salvo como una figura equívoca del habla, decir, como hace Le Doux, que es «posible que el cerebro sepa que algo es bueno o malo antes de que sepa exactamente qué es».
>
> (pág. 152)

¿Pero quién es el que se confunde? No LeDoux, ni los lectores de LeDoux, *si* leen con detenimiento, ya que se darán cuenta de que ha dado con un modo muy bueno de formular la sorprendente tesis de que un circuito especialista del cerebro puede discernir algo como, por ejemplo, peligroso o deseable, en base a una especie de rápido *triaje** que concluye *antes* de que la información pase a esas redes que completan la identificación del estímulo. (Sí, sí, ya lo sé. Sólo una persona —un médico, una enfermera o alguien por el estilo— puede llevar a cabo la conducta que llamamos *triaje*; estoy hablando en sentido «metonímico». Acostúmbrese el lector.)

En resumen, lo que estoy diciendo a mis colegas de las neurociencias es que no hay aquí caso alguno del que defenderse. Los autores sostienen que casi todos los que se dedican a la neurociencia cognitiva cometen un disparate conceptual más bien simple. Yo digo que hay que desestimar todos los cargos hasta que los autores presenten pruebas dignas de consideración. ¿Ofrecen los autores algo más que pudiera tener algún valor para las neurociencias? No ofrecen ninguna teoría ni ningún modelo positivos, ni ninguna sugerencia sobre cómo se podrían construir tales teorías o modelos, ya que, evidentemente, ése no sería el campo de la filosofía. Sus interpretaciones «correctas» de la comisurotomía y de la visión ciega, por ejemplo, consisten en insulsas reformulaciones de los fenómenos expuestos, y en

* Sistema médico de selección y clasificación de pacientes en emergencias o desastres de acuerdo con las posibilidades de supervivencia, las necesidades terapéuticas y los recursos disponibles. *(N. del t.)*

absoluto en explicaciones. Tienen razón hasta cierto punto: así es cómo aparecen estos notables fenómenos. Ahora bien, ¿cómo debemos explicarlos? La explicación debe detenerse en algún punto, como dijo Wittgenstein, pero no aquí. Bennett y Hacker citan consternados algunas de las observaciones marcadamente despectivas hacia la filosofía por parte de Glynn, Crick, Edelman, Zeki y otros (págs. 396-398). Vista la solidez de esta exposición, es comprensible que los neurocientíficos sientan tan poco interés.[23]

SITUAR DE NUEVO LA CONCIENCIA EN EL CEREBRO

Réplica a Philosophical Foundations of Neuroscience, *de Bennett y Hacker*

JOHN SEARLE

Éste es un libro extenso, de más de 450 páginas, y trata una enorme cantidad de cuestiones. Contiene muchas objeciones a mis ideas, así como un apéndice dedicado específicamente a criticarme. Voy a limitar mis observaciones a ciertas cuestiones centrales del libro, y a responder a las que considero son las críticas más importantes de Bennett y Hacker. Pero no intentaré discutir todas las cuestiones de enjundia que se plantean en su libro.

Dado que la mayoría de mis comentarios serán críticos, quiero empezar por señalar algunos puntos de acuerdo importantes. Los autores aciertan al señalar que en la percepción solemos percibir objetos reales del mundo, y no reproducciones o imágenes interiores de objetos. También están en lo cierto al observar que la relación normal que uno mantiene con sus propias experiencias no es epistémica. Es un error pensar en nuestra relación con nuestras percepciones basándonos en el modelo del «acceso privilegiado» u otro tipo cualquiera de «acceso» epistémico, y por su parte el modelo de la «introspección», por el que *spectamus intro* (contemplamos por dentro) nuestra propia mente, es irremediablemente confuso. He señalado todos estos puntos en diversos escritos, y me complace ver que estamos de acuerdo en estas cuestiones.[1]

Les remití el primer borrador de este artículo y ellos me señalaron amablemente los pasajes en los que, en su opinión, había interpretado o formulado mal sus ideas, lo cual me permitió corregir determinados malentendidos.

Su concepción básica de los fenómenos mentales es a mi entender errónea, y en lo que sigue intentaré explicar por qué. Con el fin de enunciar la diferencia entre mis ideas y las suyas, empezaré con un breve sumario de algunas de mis concepciones, para a continuación exponer con claridad las suyas por contraste. En aras de la brevedad, limitaré mi discusión a la conciencia, aunque es posible hacer, *mutatis mutandis*, observaciones similares sobre la intencionalidad.

La conciencia como fenómeno biológico

1. Por definición, la conciencia consiste en estados (usaré «estados» para denotar estados, procesos, acontecimientos, etc.) que son *cualitativos* y *subjetivos*. Patologías aparte, los estados conscientes sólo se dan como parte de un *único campo unificado de conciencia*. La conciencia es cualitativa en el sentido de que para cada estado consciente hay un determinado carácter cualitativo, un aspecto de cómo-es o cómo-se-siente. Por ejemplo, el carácter cualitativo de beber cerveza es diferente del carácter cualitativo de escuchar la *Novena Sinfonía* de Beethoven. Estos estados son *subjetivos* en el sentido ontológico de que únicamente existen en tanto que son experimentados por un sujeto humano o animal. Y son *unificados* en el sentido de que cualquier estado consciente, como la presente sensación del teclado bajo mis dedos, existe como parte de un gran estado consciente, mi campo de conciencia presente. Dado su carácter subjetivo y cualita-

tivo, a estos estados a veces se les llama «*qualia*». En general, no me parece un concepto útil porque implica una distinción entre aquellos estados conscientes que son cualitativos y aquellos que no lo son, y desde mi punto de vista no existe tal distinción. «Conciencia» y «*qualia*» son simplemente términos de igual extensión. Sin embargo, dado que Bennett y Hacker niegan la existencia de los *qualia*, en este texto voy a usar el término para resaltar los puntos de desacuerdo. Cuando digo que los estados conscientes existen, quiero decir que los *qualia* existen. Cuando ellos dicen que los estados conscientes existen, quieren decir algo totalmente distinto, como veremos.

Los fenómenos conscientes son fenómenos concretos que ocurren en el espacio-tiempo. No son entes abstractos como los números. Antes creía que la cualitatividad, la subjetividad y la unidad eran tres características distintivas de la conciencia, pero, después de reflexionar sobre ello, me parece claro que cada una implica la siguiente. Son tres aspectos diferentes de la esencia de la conciencia: la subjetividad cualitativa y unificada.

2. Estos estados, los *qualia*, están causados enteramente por procesos cerebrales. No estamos muy seguros de cuáles son los mecanismos causales, pero parece que la excitación de las neuronas en las sinapsis desempeña un papel funcional especialmente importante.

3. Los estados conscientes existen en el cerebro. Se hacen realidad en el cerebro como características de alto nivel o del sistema en su conjunto. Por ejemplo, los pensamientos conscientes sobre nuestras abuelas son procesos que se producen en el cerebro, pero, por lo que sabemos, ninguna neurona en concreto puede causar ni realizar pensamientos sobre una abuela. La conciencia es una característica del cerebro en un nivel superior al de las neuronas individuales.[2]

Evidentemente, hay mucho más que decir sobre la conciencia, y yo he dicho parte de ello en otros lugares. Considero que los principios expuestos en 1, 2 y 3 son verdades de sentido común científico más o menos culto. Pero el interés de la presente discusión radica en que, por increíble que parezca, *Bennett y Hacker niegan los tres principios.*

Sostienen haber demostrado que la noción de *qualia*, la noción del carácter cualitativo de las experiencias conscientes, es «incoherente». También dicen que: «diferimos de Searle [...] cuando dice que los fenómenos mentales son causados por procesos neurofisiológicos del cerebro, y son ellos mismos rasgos del cerebro» (pág. 446).

Aunque sólo sea para hacernos una idea de la magnitud de su impugnación, y sus consecuencias para la filosofía y la neurociencia, apliquemos mis tres principios a un ejemplo real. En estos momentos veo una mano enfrente de mi cara. ¿Cuáles son las partes componentes del acontecimiento de mi visión de la mano? Bueno, en primer lugar, tiene que haber ahí una mano, y debe haber un determinado impacto en mi aparato visual y neurobiológico (le ahorro los detalles al lector). En el caso normal, como el de la visión no ciega, este impacto producirá una experiencia visual consciente, en mi sentido de acontecimiento cualitativo y subjetivo, un *quale*. Quiero subrayar que la experiencia visual tiene todas las características que acabo de mencionar: es cualitativa, subjetiva y existe como parte de un campo unificado. Está causada por procesos cerebrales y existe en el cerebro. Así pues, tenemos tres componentes de la escena visual: el que percibe, el objeto percibido y la experiencia visual cualitativa. Gran parte de los mejores trabajos en neurociencia constituyen un esfuerzo por explicar de qué forma los procesos cerebrales causan la experiencia visual y dónde y cómo se produce ésta en el cerebro. Asombrosamente, Bennett y Hacker niegan la existencia de la experiencia visual en este sentido, en el sentido de *quale*. Es acertado

señalar, como ellos hacen, que lo que yo percibo es una mano y no una experiencia visual, pero negar rotundamente que hay una experiencia visual cualitativa, en el sentido de un *quale* visual, es cuando menos estrafalario. ¿Qué ocurre cuando cierro los ojos, por ejemplo? La experiencia visual cualitativa se acaba. Por esto dejo de ver la mano, porque dejo de tener la experiencia visual. Observemos que la presencia de la mano es esencial para que yo realmente vea la mano, pero no es esencial para la existencia de la experiencia visual porque, en el caso de las alucinaciones, puedo tener una experiencia indistinguible sin que haya mano alguna.

Bennett y Hacker no son los primeros autores que niegan la existencia de los *qualia*, pero su negación no la motiva el materialismo insensato de quienes temen que si aceptan la existencia de los fenómenos mentales irreductiblemente subjetivos se van a encontrar a Descartes como compañero de cama. Entonces ¿qué motiva, qué *podría* motivar, una negación de la existencia de estados conscientes, tal como los he definido?

La visión Wittgensteiniana

La mejor forma de entender el libro de Bennett y Hacker es verlo como una aplicación de la filosofía de la mente de Wittgenstein a la neurociencia actual. Gran parte de la originalidad del libro está en el hecho de que esto no se ha hecho nunca antes. La postura de Bennett y Hacker es, por lo que yo sé, única en los debates contemporáneos de filosofía de la mente.

¿Cuál es, entonces, (su interpretación de) la filosofía de la mente de Wittgenstein, y cómo la aplican a la neurociencia? Un pasaje clave de Wittgenstein, que ellos citan, es:

«Sólo del ser humano y de lo que se parece a un ser humano (se comporta como tal) se puede decir: tiene sensaciones; ve, es ciego; oye, es sordo; es consciente o inconsciente» (pág. 71).

Es decir, sólo *tiene sentido* atribuir predicados mentales a algo que sea un ser humano o se comporte como tal. ¿Y cuál es el papel de la conducta en la adscripción de estos predicados? La conducta proporciona no sólo las bases de la inducción de la presencia de fenómenos mentales, sino también los criterios lógicos. Deberíamos entender la manifestación de la conductua como el *criterio* lógico para la aplicación de estos conceptos. Hay una conexión de significado entre la conducta externa y los conceptos mentales, porque sólo de un ser capaz de exhibir una determinada forma de conducta podemos decir que tiene un fenómeno mental. Dos frases clave son las siguientes (pág. 83): «*Los criterios de la atribución de un predicado psicológico son en parte constitutivos del significado de ese predicado*» y, en la misma página: «*El cerebro no cumple los criterios para ser un posible sujeto de predicados psicológicos*».

Ahora bien, puesto que Bennett y Hacker aceptan esta concepción wittgensteiniana, creen que su inmediata consecuencia lógica es que la conciencia no puede existir en el cerebro, y que los cerebros no pueden realizar actividades mentales, como las de pensar o percibir, porque son incapaces de exhibir la conducta adecuada (pág. 83). Sólo de la persona en su totalidad o, en el caso de los animales, del animal en su totalidad, podemos decir que padecen dolor o están enfadados, porque sólo el animal en su totalidad es capaz de exhibir la conducta constitutiva de parte de las condiciones de aplicación del concepto en cuestión. Dada la conexión lógica entre los estados mentales y la conducta, no podemos hacer la tradicional separación entre el fenómeno mental y su manifestación externa. La conexión lógica explica por qué podemos literalmente ver que alguien

está enfadado o sufre algún dolor o es consciente o inconsciente. Además, Bennett y Hacker están obligados ahora a negar la existencia de los *qualia*, ya que éstos, si existieran, existirían en cerebros, y esto entra en contradicción con la tesis de que la conciencia no puede existir en cerebros. Para ellos, que la conciencia existe en cerebros no sólo es falso: no tiene sentido. Es como decir que la conciencia existe en números primos.

También hacen numerosos movimientos wittgensteinianos acerca de cómo aprende el niño el vocabulario mental y por qué tenemos un vocabulario mental. Por lo que recuerdo, no emplean la idea de un «juego del lenguaje», pero está implícita a lo largo de todo el libro. El juego del lenguaje al que jugamos con las palabras mentales exige para su aplicación unos criterios conductuales observables.

Ésta es su visión, y creo que la mayor parte de las tesis sustantivas del libro en realidad se derivan de ella. ¿Es una base válida desde la que criticar los estudios contemporáneos? ¿Es suficiente para refutar la idea de que la conciencia consiste en una subjetividad unificada y cualitativa, causada por procesos cerebrales y realizada en cerebros? No lo creo, por las razones que siguen. Supongamos que están en lo cierto en lo que se refiere a la base lógica, a la necesidad de una conducta observable públicamente para poder jugar al juego del lenguaje, a la imposibilidad de un lenguaje privado, y todo lo demás. ¿Qué se sigue de ello? No se sigue ninguna de sus espectaculares conclusiones. Una vez que aceptamos, como debemos hacer, y como hizo Wittgenstein, que hay una distinción entre el dolor y la conducta de dolor, entre los sentimientos de ira y la conducta de ira, entre el pensamiento y la conducta de pensamiento, etc., entonces podemos centrar la atención de nuestra investigación neurobiológica en el dolor, los sentimientos de ira, la experiencia visual, etc., y

olvidarnos de la conducta. Del mismo modo que los antiguos conductistas confundían las pruebas conductuales de los estados mentales con la existencia de los propios estados mentales, los seguidores de Wittgenstein cometen un error más sutil, pero fundamentalmente similar, cuando confunden los criterios de la aplicación de conceptos mentales con los propios estados mentales. Es decir, confunden los criterios conductuales para la *adscripción* de los predicados psicológicos con los *hechos adscritos por estos* predicados psicológicos, lo cual es un grave error.

Supongamos que Wittgenstein tiene razón cuando dice que no podríamos tener el vocabulario relativo al dolor a menos que haya unas formas de conducta de dolor que se puedan expresar públicamente. Del mismo modo, si me pregunto: «¿Qué hecho relativo a mí hace que sea el caso que yo sienta dolor?», no encuentro ningún hecho relativo a mi conducta que hace que sea el caso que yo sienta dolor. El hecho relativo a mí que hace que sea el caso que yo sienta dolor es el hecho de que tengo un determinado tipo de sensación desagradable. Y lo que vale para el dolor vale para la ira, el pensamiento y todo lo demás. Aun en el caso de que el planteamiento wittgensteiniano sea correcto al cien por cien como análisis filosófico del funcionamiento del *vocabulario*, igualmente podemos inferir siempre, en cualquier caso individual, la existencia del sentimiento interior, cualitativo y subjetivo a partir de su manifestación en la conducta exterior. Bennett y Hacker señalan que, aunque se pueda establecer la distinción en casos individuales, no podría ser el caso de que nunca hubiera ninguna manifestación de dolor públicamente observable, ya que, de otro modo, no podríamos usar el vocabulario relativo al dolor. Supongamos que tienen razón en este punto. De todos modos, cuando investigamos la ontología del dolor —no las condiciones para jugar al juego del lenguaje, sino la propia ontología del fenómeno en sí—

podemos olvidarnos de la conducta exterior y limitarnos a averiguar de qué forma el cerebro causa las sensaciones internas.

Obsérvese que en el pasaje citado, Wittgenstein habla de lo que podemos *decir*:

«Sólo del ser humano y de lo que se parece a un ser humano (se comporta como tal) se puede *decir*: tiene sensaciones; ve, es ciego; oye, es sordo; es consciente o inconsciente» (pág. 71; la cursiva es mía).

Pero supongamos que eliminamos de este pasaje la palabra «decir» y lo escribimos así:

«Sólo los seres humanos y las cosas que se comportan como seres humanos pueden realmente *ser* conscientes». Si tomamos esto como una afirmación conceptual o lógica, es obviamente falsa. Supongamos que la tomamos como:

«Es una verdad conceptual o lógica que sólo las cosas que se comportan como seres humanos pueden realmente ser conscientes».

Pero esto, como verdad conceptual o lógica, parece ser simplemente errónea. Por ejemplo, los moluscos y los crustáceos, como las ostras y los cangrejos, no se comportan en absoluto «como seres humanos», pero este hecho en sí mismo no resuelve la cuestión de si son conscientes o no. Con independencia de hasta qué punto su conducta difiere de la conducta humana, las ostras seguirían siendo conscientes si tuvieran el tipo adecuado de procesos neurológicos en su sistema nervioso. Supongamos que disponemos de una ciencia perfecta del cerebro y sabemos exactamente cómo se produce la conciencia en los seres humanos y en los animales superiores. Si descubriéramos entonces que el mecanismo productor de la conciencia está presente en las ostras pero no en los caracoles, tendríamos buenas razones, sin duda unas razones aplastantes, para suponer que las ostras son conscientes y los caracoles probablemente no. La existencia misma de la conciencia no tiene nada que ver con la conducta si

bien, en los seres humanos, la existencia de conducta es (un criterio) esencial para el funcionamiento del juego del lenguaje. La pregunta «¿cuáles de los animales inferiores son conscientes?» no se puede resolver mediante el análisis lingüístico.

Wittgenstein ofrece una explicación general del funcionamiento del vocabulario mental. Señala que es un error interpretar el funcionamiento del juego del lenguaje desde el modelo de pruebas inductivas externas de la presencia de fenómenos privados internos. «Un proceso interno requiere un criterio externo», nos recuerda. Pero, incluso si aceptamos esta explicación del vocabulario, nada nos impide dar una versión neurobiológica de cómo los estados de conciencia son causados por procesos cerebrales y realizados en sistemas del cerebro. Además, la exigencia de que el sistema, la persona en su totalidad, sea capaz de manifestar una conducta no implica que no pueda haber un elemento del sistema, el cerebro, que sea el lugar de los procesos conscientes. Éste es otro tema, que explicaré con mayor detalle en el apartado siguiente.

La falacia, en pocas palabras, reside en confundir las reglas de uso de las palabras con la ontología. Del mismo modo que el antiguo conductismo confundía las *pruebas* de estados mentales con la *ontología* de los estados mentales, este conductismo lógico wittgensteiniano confunde *las bases para hacer la atribución* con el *hecho que se atribuye*. Es una falacia decir que las condiciones para el correcto funcionamiento del juego del lenguaje son condiciones para la existencia de los fenómenos en cuestión. Supongamos que tenemos una ciencia perfecta del cerebro, de modo que sabemos de qué modo el cerebro produce el dolor. Supongamos que construimos una máquina que fuera capaz de tener conciencia e incluso capaz de tener conciencia de sentir dolores. Podríamos diseñar una máquina de forma que no mostrara ningún tipo de conducta de dolor. De-

pendería de nosotros. Hay, sin duda, formas de enfermedades reales en las que las personas padecen dolores sin una conducta de dolor. En algunos casos del síndrome de Guillain-Barré, el paciente es completamente consciente pero está paralizado por completo, totalmente incapaz de exhibir una conducta que se corresponda con sus estados mentales. Bennett y Hacker señalan que éste no podría ser el caso para todos los dolores, que nadie que sufra dolor mostrara jamás una conducta de dolor, porque entonces no podríamos aplicar las palabras. Incluso si eso es correcto, es una condición para el adecuado funcionamiento del juego del lenguaje, no una condición para la existencia de dolores.

Resumen de la argumentación expuesta hasta ahora

Creo que una vez que se elimina esta falacia básica, el razonamiento central del libro se desmorona. Hablaré de los detallados argumentos de nuestros autores en un momento, pero antes quiero resumir los argumentos expuestos hasta aquí. Wittgenstein sostiene que una condición de la posibilidad de un lenguaje para describir los fenómenos mentales internos son las manifestaciones públicamente observables de esos fenómenos. La conducta no es simplemente una prueba inductiva, sino que es un criterio para la aplicación de los conceptos. Supongamos, en aras de la argumentación, que tiene razón. Bennett y Hacker creen que de ello se sigue que los fenómenos mentales no podrían existir en los cerebros, porque éstos no pueden mostrar la conducta que sirve de criterio. Pero esto no se sigue de que la conducta observable sea un criterio de aplicación del lenguaje sobre lo mental. Todo lo que se podría seguir de ello es que si podemos hablar de estados mentales en el cerebro, entonces éste debe

formar parte de un mecanismo causal capaz de producir conducta (me ocuparé de este punto con mayor detalle en el apartado siguiente). Y normalmente lo es. Pero incluso en los casos en que no lo es, debemos distinguir la propia existencia de los fenómenos mentales de la posibilidad de hablar sobre ella. El hecho que hace que sea el caso que yo tenga dolor cuando realmente tengo dolor es la existencia de un determinado tipo de sensación. El hecho de que yo manifieste o no esa sensación en la conducta es irrelevante para su propia existencia.

El argumento principal del libro: La falacia mereológica

El argumento más importante del libro, el que se repite una y otra vez, es el desenmascaramiento de lo que Bennett y Hacker llaman la «falacia mereológica», que ellos definen como la atribución a las partes de aquello que sólo tiene sentido cuando se atribuye al todo. La forma típica que esta falacia adopta, según ellos, es que los neurocientíficos dicen cosas como que el *cerebro* piensa, percibe, confía, se pregunta, decide, etc., cuando de hecho la caracterización correcta sería que *la persona entera* piensa, percibe, confía, se pregunta, decide, etc. La falacia está en atribuir a la parte, el cerebro, lo que sólo tiene sentido cuando se atribuye al todo, a la persona en su totalidad. Espero que sea evidente que esto se sigue de la visión wittgensteiniana: dado que la parte, es decir, el cerebro, no puede exhibir la conducta consciente, y dado que ésta es esencial para la atribución de la conciencia, no podemos atribuir dolores al cerebro.

Voy a hacer una observación pedante, cuya importancia se percibirá más adelante. Según la propia explicación de nuestros autores, éste

no puede ser, estrictamente hablando, un caso de falacia mereológica, la falacia de atribuir a la parte lo que sólo tiene sentido cuando se atribuye al todo, porque si lo fuera, podríamos eliminar la falacia simplemente con la adición de una referencia al resto del cuerpo. La relación del cerebro con el resto del *cuerpo* es sin duda una relación de parte y todo. El cerebro es una parte de mi cuerpo. Bennett y Hacker dicen que únicamente la *persona* puede ser el sujeto de atribuciones psicológicas, no el cerebro solo. Pero la persona no está relacionada con el cerebro como el todo lo está con la parte. Esto no implica que la persona sea algo distinto del cuerpo ni que esté «por encima» de él. Lamentablemente los autores nunca nos dicen qué es una persona, y sin embargo creo que es algo crucial para la exposición en su conjunto, de hecho para la entera discusión. Lo que ellos llaman una falacia mereológica es más bien un error categorial en el sentido de Ryle. Según la exposición de Bennet y Hacker, las personas pertenecen a una categoría lógica diferente a la de los cerebros y, por esta razón, las atribuciones psicológicas a las personas no tienen sentido atribuidas al cerebro. Retomaré también este punto más adelante.

Como bien saben nuestros autores, de hecho hay (al menos) tres tipos diferentes de atribuciones subpersonales de fenómenos psicológicos, y los argumentos que usan en contra de uno de ellos no son necesariamente aplicables a los otros. En primer lugar, está el cerebro como sujeto y agente (por ejemplo: «El cerebro piensa»). En segundo lugar, está el cerebro como ubicación de los procesos psicológicos (por ejemplo: «El pensamiento tiene lugar en el cerebro»). Y, en tercer lugar, están los microelementos como agentes (por ejemplo: «Las neuronas individuales piensan»). Veamos en orden cada uno de ellos.

Primero, el cerebro como sujeto y agente. Como observaba antes, es bastante habitual en la literatura tanto filosófica como neurobioló-

gica describir la cognición usando para ello verbos cognitivos cuyo sujeto es «el cerebro». Así, es algo común decir que el cerebro percibe, el cerebro piensa, el cerebro decide, etc., algo que Bennett y Hacker consideran inaceptable por la razón que he expuesto: que el cerebro no puede exhibir la conducta apropiada. En el habla corriente tenemos que decir que la persona decide. Soy yo quien decide votar por el candidato demócrata, no mi cerebro.

El argumento para esta conclusión, como decía antes, es que el cerebro, dado que no puede exhibir conducta, no puede ser el sujeto de predicados psicológicos. Pero una vez que nos percatamos de lo endeble de tal argumento, ¿podemos pensar en cualquier otra razón para rechazar la atribución de procesos psicológicos al cerebro? Estoy de acuerdo con nuestros autores en que es extraño decir, por ejemplo: «Mi cerebro decidió votar a los demócratas en las últimas elecciones». ¿Por qué? Volveré a este tema más adelante.

Segundo, el cerebro como lugar. Una segunda forma de atribución, completamente distinta de la primera, es enunciar *dónde* se producen los procesos y los acontecimientos psicológicos. Y aquí la tesis es que tienen lugar en el cerebro. Bennett y Hacker son conscientes de la distinción entre tratar el cerebro como el agente y sujeto de los procesos psicológicos y tratarlo como el lugar de los procesos psicológicos; pero se oponen a ambas ideas. Creen tanto que el cerebro no puede pensar como que el pensamiento no puede tener lugar en el cerebro. Pero para demostrar que el cerebro no puede ser el lugar de tales procesos precisarían un argumento independiente, un argumento que no he podido encontrar. Con rigor, aunque el argumento wittgensteiniano fuera válido, no se podría esgrimir contra todas estas atribuciones. ¿Por qué no? El argumento dice que el agente de un proceso psicológico debe ser un sistema capaz de exhibir la conducta adecuada. Así, en nuestro ejemplo de la visión, un sistema

capaz de ver debe ser capaz de exhibir la conducta apropiada. De modo que no podemos decir de un cerebro que ve, sólo podemos decir esto del sistema entero, es decir, de la persona. Pero esto no nos impide identificar la experiencia visual como un componente de la visión ni localizar la experiencia visual en el cerebro. Todo lo que el argumento wittgensteiniano exige es que el cerebro sea parte de un mecanismo causal de un sistema global capaz de producir la conducta. Y esta condición se puede cumplir incluso cuando determinados procesos psicológicos se localizan en el cerebro.

Para entenderlo, consideremos una analogía. Supongamos que alguien dice: «No deberíamos decir que el estómago y el resto del tracto digestivo digieren los alimentos. Sólo la persona en su conjunto puede digerir los alimentos». En cierto sentido, así es. Pero observemos que, para los propósitos de la investigación acerca de cómo funciona la digestión, podemos preguntarnos dónde y cómo se producen los procesos digestivos concretos. Y la respuesta es que tienen lugar en el estómago y el resto del tracto digestivo. Ahora bien, del mismo modo alguien podría insistir en que soy yo, la persona, quien percibe y piensa conscientemente, y no mi cerebro. Asimismo, uno podría preguntar a continuación en qué parte de la anatomía tienen lugar los procesos conscientes, y la respuesta obvia es que tienen lugar en el cerebro. Bennett y Hacker son conscientes de la distinción, pero no parecen darse cuenta de que la afirmación de que el cerebro no puede ser el lugar de los procesos psicológicos exigiría una argumentación distinta, y no encuentro tal argumentación en su libro. Dicen ellos: «La ubicación *del acontecimiento de que una persona piensa un determinado pensamiento* es el lugar donde la persona está cuando se le ocurre el pensamiento» (pág. 180). Así es, sin duda, pero ello no implica que mis pensamientos no ocurran también en mi cabeza. Determinados pensamientos se me ocurren en este pre-

ciso momento. ¿Dónde? En esta habitación. ¿Dónde exactamente de esta habitación? En mi cerebro. De hecho, con el desarrollo de la resonancia magnética funcional (RMf) y otras técnicas de imaginería, nos estamos acercando cada vez más a la posibilidad de decir en qué lugar exacto del cerebro tienen lugar los pensamientos.

La ubicación de los estados conscientes y los procesos cerebrales como causa de esos estados

La concepción que he anticipado es que los procesos mentales conscientes tienen lugar en el cerebro y están causados por procesos neuronales de nivel inferior. ¿Cuál es la concepción de Bennett y Hacker? En tanto que niegan que los procesos mentales tienen lugar en el cerebro, creo que son incapaces de dar una explicación coherente de la ubicación y la causa de la conciencia. Piensan que los procesos neurales son una condición necesaria de la conciencia, pero no enuncian, y en mi opinión no pueden enunciar, la tesis obvia de que, en las circunstancias adecuadas, los procesos neurobiológicos son la causa suficiente de la conciencia. Mis estados conscientes actuales, todos ellos *qualia*, son causados por procesos neuronales de nivel inferior en el cerebro.

Creo que algunas de las argumentaciones más endebles del libro se refieren a este tema de la ubicación y la causa de los estados conscientes. Dicen nuestros autores: «No existe *tal cosa* como un proceso mental (como el de recitar el alfabeto en la imaginación) que ocurre en una parte del animal, sea esta parte los riñones o el cerebro» (pág. 112). «Lo que pasa *en* el cerebro son procesos neurales, que tienen que ocurrir para que la persona, el poseedor del cerebro, pase *por* los procesos mentales relevantes» (pág. 112).

Creo que este pasaje contiene un profundo error, que voy a examinar paso a paso.

Supongamos que recito el alfabeto en silencio, «en mi cabeza», como diríamos. Es un acontecimiento real del mundo real. Al igual que todos los acontecimientos reales, tiene lugar en el espacio-tiempo. Entonces, ¿dónde ha ocurrido? Bennett y Hacker dicen que yo, la persona, *pasé por* el proceso mental. Algo indudablemente cierto, pero ¿dónde exactamente del espacio-tiempo tuvo lugar el acontecimiento mental consciente, temporal y espacialmente situado, de mi recitación silenciosa del alfabeto? Nuestros autores no pueden responder a esta pregunta, excepto diciendo tal vez que ocurrió en Nueva York, en esta habitación o cosas por el estilo. Pero con esto no basta. Creo que es evidente que dentro de mi cerebro ocurrió un acontecimiento consciente, un conjunto de *qualia*. Y, como ya he dicho antes, con el avance de técnicas de imaginería como la RMf estamos mucho más cerca de poder decir dónde ocurrió exactamente.

Su negación de la realidad y ubicación espacial de los *qualia* impide a nuestros autores dar una explicación *causal* coherente de la relación de los procesos neurales con los acontecimientos mentales. Dicen que los procesos neurales son causalmente *necesarios* («tienen que ocurrir») para que ocurra un proceso mental. Pero, en este contexto, necesitamos saber qué es causalmente *suficiente*, qué fue lo que hizo que yo «pasara» por un proceso mental. Y siempre que hablamos de causas, debemos decir exactamente qué causa exactamente qué. En mi explicación, el acontecimiento mental consciente es causado en su totalidad por el cerebro y se realiza en el cerebro. En este contexto, la excitación de ciertas neuronas es causalmente suficiente para producir esos *qualia*, esos acontecimientos mentales conscientes. ¿Qué dicen ellos? No pueden decir que la excitación de las

neuronas haya causado los *qualia*, la experiencia cualitativa, porque han negado la existencia de los *qualia*. No pueden decir que la excitación de las neuronas haya causado la conducta externa, porque no ha habido conducta externa alguna. Entonces, ¿cuál es exactamente la naturaleza del proceso mental por el que yo pasé y cuáles son exactamente las causas que hicieron que se produjera? Las neuronas se activaron... ¿y luego qué? Las neuronas deben activarse para que yo «pase» por los procesos mentales. ¿Pero en qué consiste pasar por los procesos mentales si no hay *qualia* ni conducta? Bennet y Hacker no tienen respuesta para estas preguntas y, considerando su teoría en su conjunto, no creo que la puedan tener.

Desde el supuesto wittgensteiniano, si el cerebro no puede exhibir una conducta, entonces no puede ser el sujeto ni el agente de atribuciones mentales. Creo que esto es un error. Pero, con independencia de que lo sea o no, es preciso distinguir entre el argumento de que el cerebro no puede ser el *sujeto* de verbos psicológicos y el argumento de que el cerebro no puede ser el *lugar* de procesos psicológicos. Y Bennett y Hacker no distinguen ambos argumentos. Son conscientes de la distinción entre el cerebro como sujeto y el cerebro como lugar y niegan ambas cosas, que el cerebro pueda ser el sujeto y que pueda ser el lugar. Pero no ofrecen un argumento independiente en contra de la afirmación de que el cerebro es el lugar de, por ejemplo, los procesos del pensamiento. Supongamos que estuviéramos de acuerdo en que suena extraño decir: «Mi cerebro piensa...» Así y todo, aún puede haber procesos de pensamiento activos en mi cerebro cuando *yo pienso* . Lo máximo que el argumento wittgensteiniano puede establecer es que no debemos pensar en el cerebro como sujeto o agente. Pero de ello no se sigue que no sea el *lugar* de los correspondientes procesos. Su argumento contra el *cerebro como sujeto* no se extiende al *cerebro como lugar*.

La pregunta «¿dónde ocurren los procesos conscientes?» no es filosóficamente más abstrusa que la pregunta «¿dónde ocurre el proceso digestivo?». Los procesos cognitivos son procesos biológicos reales en la misma medida en que lo es la digestión. Y la respuesta a ambas preguntas es evidente. La digestión ocurre en el estómago y el resto del tracto digestivo; la conciencia ocurre en el cerebro y quizás otras partes del sistema nervioso central.

Atribuciones de estados psicológicos: metafóricas y literales, relativas al observador e independientes del observador

Tercero, las neuronas como sujetos y agentes. Una tercera forma de atribución que Bennett y Hacker objetan es la de atribuir procesos psicológicos a partes del cerebro. Así, por ejemplo, citan a Blakemore cuando dice que las neuronas perciben, deciden, hacen inferencias, etc. Creen que éste es también un ejemplo de la falacia mereológica.

Entiendo que interpretado oportunamente esto no es más que una metáfora inocua, o por lo menos puede serlo. En efecto, en la literatura científica se hacen atribuciones de este tipo al estómago. Se dice que el estómago sabe cuándo son precisas ciertas sustancias químicas para digerir determinados tipos de hidratos de carbono que le llegan. A mi parecer estas metáforas son o pueden ser inocuas, siempre y cuando tengamos clara la distinción entre el sentido literal e independiente del observador en el que infiero o recibo información, y los sentidos metafóricos y relativos al observador en los que decimos que mis neuronas hacen tales o cuales inferencias, o perciben tales o cuales fenómenos. Es más fácil cometer el error de confundir los

sentidos reales independientes del observador con los sentidos metafóricos relativos al observador cuando se trata del cerebro que con respecto a otros órganos, por la razón obvia de que los procesos psicológicos intrínsecos, independientes del observador, tienen lugar en el cerebro y no en el estómago ni en el resto del tracto digestivo. Estoy convencido de que Bennett y Hacker aciertan al pensar que algunos de los autores que critican no tienen una idea clara de la distinción entre las atribuciones de estos fenómenos independientes del observador y las relativas al observador. Desde una perspectiva semántica, el responsable principal es la noción de «información». Una confusión adicional proviene del uso que los ingenieros hacen de «información» en el sentido de la teoría de la información, que nada tiene que ver con la información en el sentido de que tengo información de que ocurre esto o aquello. Se nos dice, por ejemplo, que el cerebro procesa la información. Bueno, en cierto sentido, así es. Obtengo información a través de la percepción y pienso sobre ella, y luego derivo nueva información mediante inferencias. El problema es que en el cerebro (en el núcleo geniculado lateral, por ejemplo) tiene lugar todo tipo de procesos subpersonales, que se pueden describir *como si* fueran casos de pensamiento sobre la información, pero, por supuesto, ahí no hay literalmente información. No hay más que excitaciones de neuronas, que se traducen en información de un tipo consciente al final del proceso, pero en sí mismas carecen de contenido semántico. Bennett y Hacker hablan con claridad de la distinción entre el sentido de «información» en la teoría de la información y el sentido intencional. Se refieren a ellos como el sentido «ingenieril» y el sentido «semántico», respectivamente, pero no encuentro en su libro una afirmación clara de la distinción entre el sentido de «información» independiente del observador y el dependiente del observador. Yo tengo información independiente del

observador sobre mi número de teléfono. La guía telefónica tiene información dependiente del observador sobre ese mismo número de teléfono. No tengo reparos en hablar de información y procesamiento de información en el cerebro, siempre y cuando estén claras estas distinciones.

Para resumir dónde nos encontramos en este momento, he planteado tres objeciones principales a los argumentos de Bennett y Hacker. Primera, la tesis wittgensteiniana de que la conducta es un criterio para la adscripción de fenómenos mentales en un lenguaje público, aunque sea cierta, no refuta la concepción de que la conciencia puede existir en el cerebro. Segunda, una vez que distinguimos entre cerebro como sujeto y cerebro como lugar, nos percatamos de que Bennett y Hacker no disponen de un argumento independiente contra el cerebro como lugar. De hecho, por lo que sabemos, es verdad que la ubicación de todos nuestros procesos conscientes está en el cerebro. Tercera, la atribución de estados psicológicos a entes subpersonales como las neuronas puede ser inocua siempre y cuando esté claro que es un uso metafórico. Si se distingue lo literal de lo metafórico, y lo relativo al observador de lo independiente del observador, no hay nada necesariamente erróneo en tales atribuciones.

Los argumentos contra los *qualia*

Para que el argumento wittgensteiniano funcione, Bennett y Hacker precisan de un argumento independiente contra los *qualia*. ¿Por qué? Bien, si los *qualia* existieran deberían tener una ubicación, y la ubicación más obvia es el cerebro. Y esto sería incoherente con el conjunto de su teoría. A continuación examinaré sus argumentos contra los *qualia*.

Como dije antes, la conciencia consiste en procesos mentales subjetivos, cualitativos y unificados, que tienen lugar dentro del cerebro humano en el interior del espacio físico del cráneo, presumiblemente localizados en su mayor parte en el sistema talamocortical. Mi tesis es que nuestros dolores, cosquilleos y picores, por ejemplo, son *subjetivos*, en el sentido de que sólo pueden existir en la medida en que los experimenta un sujeto real, que son *cualitativos* y que ocurren como *parte de un campo consciente unificado*. Bennett y Hacker piensan que disponen de argumentos en contra de mi tesis. En primer lugar, utilizando un argumento que solía usar Ryle, sostienen que la afirmación de que cualquier dolor, cosquilleo o picor sólo los puede experimentar un único sujeto no es más que una afirmación gramatical trivial sin consecuencias ontológicas. Dicen que, exactamente de la misma forma, una sonrisa debe ser siempre sonrisa de alguien, un estornudo ha de ser estornudo de alguien o, para emplear un ejemplo de Ryle, la parada de un portero en un partido debe ser la parada de alguien. En este sentido, la privacidad de los estornudos, las sonrisas y las paradas no demuestra nada ontológicamente significativo. Lo mismo dicen de los dolores, los cosquilleos y los picores. Sí, tienen que ser el dolor, el cosquilleo o el picor de alguien, pero esto no es más que un punto gramatical trivial, que no les otorga ningún estatus especial.

La respuesta a esto, implícita en lo que ya he dicho hasta ahora, es que las expresiones que se refieren a estados conscientes no sólo tienen la característica *gramatical* de exigir un sintagma nominal personal para la identificación de los casos particulares, sino más bien que la *subjetividad* de los fenómenos va unida a su propia condición de cualitativos. Y éste no es simplemente un punto gramatical. En un dolor, un cosquilleo o un picor hay una determinada sensación cualitativa, y esto es esencial para la existencia de los dolores, los cos-

quilleos y los picores. Esta sensación cualitativa forma parte de la subjetividad ontológica en cuestión, a diferencia de las características de, por ejemplo, las paradas de un portero.

Bennett y Hacker presentan argumentos independientes contra la existencia de los *qualia*, pero estos argumentos me parecen extremadamente débiles. Dicen que el olor de una lila y el de una rosa tendrán el mismo carácter cualitativo si son igualmente agradables o desagradables (págs. 275-276). Sorprendentemente, presumen que la «cualidad» es una cuestión de grados de agradabilidad o de repugnancia. Pero no se trata de esto. Los *qualia* del olor de las rosas y el olor de las lilas no están constituidos por el grado en que son agradables o desagradables. No es eso. Lo que se quiere sugerir con el nombre de *qualia* es que el *carácter* de las experiencias es diferente. Aunque los *qualia* sí poseen los rasgos característicos de la agradabilidad o la repugnancia, la esencia que los define es la sensación *cualitativa* de la experiencia. La respuesta de Bennett y Hacker a esta cuestión también me parece pobre. Afirman que si no definimos los *qualia* como una cuestión de agradabilidad o repugnancia, entonces tendremos que individualizar la experiencia en función de cuál sea su objeto: el olor de una *lila* o el olor de una *rosa*. Y añaden que identificar la experiencia por su objeto no es identificar nada subjetivo sobre la experiencia, porque las rosas tienen una existencia objetiva. De nuevo, esto me parece un error. Por supuesto que típicamente identificamos el carácter de nuestras experiencias perceptivas por sus causas, es decir, por el objeto intencional que causa que tengamos la experiencia característica. Pero uno puede tener la experiencia e individualizar la experiencia sin las causas. Si resulta que mi actos de oler la rosa y oler la lila fueran ambos alucinaciones, ello no afectaría en absoluto a las diferencias en los *qualia*. Los aspectos cualitativos del olor de una rosa y el olor de una lila siguen siendo los

mismos haya o no realmente un objeto intencional en cada caso. A menudo discriminamos olores sin tener ni idea de qué los ha causado. En pocas palabras, decir que cuando describimos los *qualia* lo hacemos típicamente desde la perspectiva de sus objetos, es decir, los fenómenos que los causan, no constituye ninguna objeción a la existencia de *qualia* con caracteres cualitativos distintivos. La noción de *qualia* se puede definir en términos de la experiencia consciente, simplemente abstrayendo la experiencia consciente de su objeto intencional. Los *qualia* consisten en las experiencias conscientes, al margen de cómo decidamos identificarlos. Merece la pena señalar que los químicos que trabajan en la elaboración de perfumes intentan sintetizar las sustancias químicas que replican los poderes causales de cosas como las rosas y las lilas. Tratan de producir unos *qualia* de tipo idéntico al de los que producen las flores de verdad.

Bennett y Hacker invocan un tercer argumento que en mi opinión es igualmente desacertado. Dicen que, típicamente, personas diferentes pueden sentir el mismo dolor o tener la misma jaqueca. Si el lector y yo acudimos a una fiesta y bebemos demasiado vino, al día siguiente tendremos la misma jaqueca o, si ambos sufrimos la misma enfermedad que produce dolores de estómago, tendremos el mismo dolor. Creo que esto demuestra una vez más que no entienden lo que los filósofos intentan decir sobre la «privacidad» de los dolores. «Privacidad» quizá sea una mala metáfora, pero esto es irrelevante aquí. La cuestión es que cuando los filósofos hablan del mismo dolor se refieren al mismo *tipo* y no al mismo *ejemplar*. Lo que nos interesa, cuando hablamos de la privacidad de los dolores, no es que diferentes personas no puedan experimentar el mismo tipo de dolor. Claro que pueden. Lo que nos interesa es que, como ejemplar, el dolor que ellos o yo experimentamos *existe sólo en tanto que lo percibe* un sujeto consciente en particular.

La ubicación de los dolores

He dicho que todos los estados conscientes existen en el cerebro. ¿Pero qué ocurre con el dolor en un pie? Sin duda está en el pie, no en el cerebro. Bennett y Hacker objetan mis ideas en esta cuestión, y puesto que esas ideas son a primera vista contrarias a la intuición, quiero dejarlas claras. Creo que si formulamos todos los hechos con claridad, las preguntas referentes a la ubicación de los dolores tendrán respuestas obvias.

En primer lugar, el espacio real, es decir, el espacio físico. Sólo hay un espacio real, físico, y todo lo que se halla en él esta relacionado espacialmente con todo lo demás. Hoy, después de Einstein, concebimos el espacio y el tiempo como un único continuo de espacio-tiempo, y las ubicaciones de las cosas se especifican en relación con un sistema coordinado. Para nuestros propósitos, el espacio está lógicamente ordenado. Consideremos la transitividad de «en»: si la silla está en la habitación y la habitación está en la casa, entonces la silla está en la casa. Todos los acontecimientos del mundo real se producen en el espacio y el tiempo físicos. Algunos de esos acontecimientos carecen de límites bien definidos —pensemos en la Gran Depresión o en la Reforma Protestante, por ejemplo—, pero, en cualquier caso, y al igual que todos los demás acontecimientos, ocurren en el espacio y el tiempo.

Pasemos ahora al espacio experimentado y fenomenológico del cuerpo. Supongamos que sufro una herida en el pie. Esto activa una secuencia de excitaciones de neuronas que suben por mi columna vertebral, pasan por el tracto dorsolateral de Lissauer, acceden a los centros del dolor de mi cerebro y, como resultado de ello, *siento un dolor en el pie*. Esta descripción es incuestionablemente la correcta, de modo que si, por ejemplo, el médico me preguntara dónde sien-

to el dolor, yo señalaría el pie y no la cabeza. Es decir, para señalar la ubicación del dolor, señalo mis miembros anatómicos en el espacio real. La pregunta es ahora: ¿cuál es exactamente la relación entre el espacio físico real y el espacio experimentado y fenomenológico del cuerpo?

Para responder esta cuestión, debemos preguntarnos cómo crea el cerebro el espacio fenomenológico del cuerpo. El cerebro crea una imagen del cuerpo, una conciencia fenomenológica de las partes del cuerpo, su estado y las relaciones entre ellas. El cerebro crea una conciencia de mi pie incluida en la imagen del cuerpo y, por consiguiente, cuando siento el dolor, una conciencia del dolor en tanto que situado en mi pie. En síntesis, el cerebro crea un espacio corporal fenomenológicamente real y un dolor en ese espacio corporal. No hay duda acerca de la realidad fenomenológica de estos fenómenos; la única pregunta es: ¿cuál es exactamente la relación entre el espacio fenomenológico del cuerpo y el espacio físico real de mi cuerpo?

Los problemas surgen cuando intentamos tratar el espacio corporal fenomenológico como si fuera idéntico al espacio físico real del cuerpo. Observemos que la transitividad de «en» no funciona si intentamos pasar del espacio fenomenológico al espacio físico. Siento un dolor en el pie y mi pie está en la habitación, pero el dolor no está en la habitación. ¿Por qué no? El enigma que supone la relación del espacio fenomenológico con el espacio físico se complica aún más cuando consideramos el dolor en miembros fantasmas. El paciente siente un dolor en el pie, pero no tiene pie. El dolor es real, ¿pero dónde está? Bennett y Hacker responden esta pregunta como sigue. «*Así pues, realmente siente un dolor donde debería estar su pie (es decir, en su miembro fantasma)*» (pág. 125). Pero un miembro fantasma no es un tipo de miembro, como un miembro herido o un miembro quemado por el sol. Los miembros fantasmas no existen

como objetos en el espacio real. De modo que si consideramos que la observación de nuestros autores se refiere literalmente al espacio físico, nos encontramos con el resultado absurdo de que el paciente «tiene un dolor donde estaría su pie», es decir, en la cama. El dolor está justamente ahí, debajo de la sábana. ¿Exactamente por qué razón es esto absurdo? Porque en el espacio físico de las camas y las sábanas no hay dolores. Los dolores sólo pueden existir en los espacios fenomenológicos del cuerpo. Y si consideramos que el paciente se refiere a su imagen corporal fenomenológica, lo que dice es cierto. El hombre siente un dolor en el pie y, aunque no tiene un pie físico, sigue teniendo un pie fantasma fenomenológico en su imagen del cuerpo.

Sin embargo, y éste es el punto crucial, el dolor es un acontecimiento real del mundo real, por lo tanto debe tener una ubicación en el espacio-tiempo real. No puede estar en el pie del hombre, porque éste no tiene pie. No puede estar donde estaría el pie si el hombre lo tuviese, porque no hay nada entre las sábanas. Evidentemente, está en su pie fantasma, pero su pie fantasma no es un objeto con una ubicación espacial como parte del cuerpo, como la de un pie real. Confío en que resulte obvio que, en el espacio físico del mundo real, la ubicación espacial del dolor fantasma del pie fantasma está en la imagen del cuerpo, que está en el cerebro. En el espacio físico real, tanto el dolor del pie real como el dolor del pie fantasma están en el cerebro, junto con el resto de la imagen corporal.

¿Es una persona un cerebro (dentro de un cuerpo)?

Voy a ocuparme ahora de la cuestión que he mencionado antes, la aparente impropiedad de atribuir actividades psicológicas al cerebro y no a la persona en su conjunto. He mostrado mi desacuerdo con los ar-

gumentos que nuestros autores emplean para demostrar que el cerebro no puede contener procesos psicológicos, pero estoy de acuerdo en que suena extraño decir, por ejemplo: «Mi cerebro ha decidido votar a los demócratas». ¿Por qué suena mal? ¿Por qué extendemos con toda naturalidad el uso de algunos predicados referidos a la persona a algún aspecto de la persona, como su cuerpo? Consideremos las cuatro frases siguientes:

1. Yo peso 75 kilos.
2. Sé distinguir el azul del violeta cuando los veo.
3. He decidido votar a los demócratas.
4. Soy propietario de un inmueble en la ciudad de Berkeley.

En el caso de 1, no vacilaríamos en sustituir «yo» por «mi cuerpo», es decir, *yo* peso 75 kilos si y sólo si *mi cuerpo* pesa 75 kilos. Ambas afirmaciones parecen ser equivalentes. Tampoco veo problema alguno en hacer una sustitución parecida en la frase 2. Mi cerebro, concretamente mi sistema visual, incluidos mis ojos, sabe distinguir el azul del violeta. Pero me parece que vacilaríamos mucho más en hacer una sustitución por el estilo en la frase 3. Si quiero expresar que he decidido votar a los demócratas, parece más bien desconcertante decir que mi cerebro o el cerebro que hay en mi cuerpo ha decidido votar a los demócratas. Una sustitución análoga en la frase 4 es, en mi opinión, aún más extraña. Decir que el cerebro que hay en mi cuerpo es propietario de un inmueble en Berkeley, o que mi cuerpo es propietario de un inmueble en Berkeley, suena manifiestamente raro. Bennett y Hacker rechazan la sustitución en las frases 2 y 3. Invocando el argumento wittgensteiniano, no admitirían ni «Mi sistema visual (en mi cerebro) sabe distinguir el azul del violeta» ni «Mi cerebro ha decidido votar a los demócratas». Más

arriba he expuesto las razones para rechazar el argumento wittgensteiniano; no obstante, concedamos que suena lógicamente impropio decir que mi cerebro ha decidido votar a los demócratas. Aun en el caso de que el argumento wittgensteiniano sea un error, debemos abordar esa impropiedad. A muchísimos neurobiólogos y filósofos les parece completamente natural atribuir actividades psicológicas al cerebro.

¿Cómo resolver esta disputa? Creo que en el nivel en que estamos debatiendo esta cuestión ahora, hay una forma bastante fácil de resolver la aparente disputa. Siempre que dispongamos de una formulación alternativa para cualquier afirmación *de que S*, una primera forma de enfocar la validez de la formulación alternativa es preguntarnos qué causa el hecho *de que p*, si *S* expresa la proposición *de que p*. No tenemos problemas con la primera frase «Yo peso 75 kilos» cuando la sustituimos por «Mi cuerpo pesa 75 kilos» porque sabemos qué hecho relativo a mí causa que pese 75 kilos, concretamente que esto es lo que pesa mi cuerpo. No tengo objeción alguna a un cambio similar para la frase 2, porque si nos preguntamos qué hecho relativo a mí causa que yo sea capaz de distinguir el azul del violeta, el hecho en cuestión es que mi sistema visual es capaz de distinguir el azul del violeta. Pero las frases 3 y 4 son muy distintas de la 1 y la 2, porque no sólo requieren la existencia y las características de un cerebro integrado en un cuerpo, sino más bien que ese cerebro esté situado en un contexto social y sea capaz de llevar a cabo acciones sociales. En el caso de 3, a diferencia de 4, podemos hacer abstracción de la situación social e identificar un componente puramente psicológico. Dada mi situación social y política, y en ese contexto, determinadas actividades que tienen lugar en mi cerebro constituyen el hecho de que haya decidido votar a los demócratas. La misma vacilación que nos impide atribuir al cerebro algo que requiera que la persona encarnada en el cuerpo esté socialmente situada, también

nos hace reticentes a atribuir al cerebro cualquier tipo de acción o de condición de agente. Así pues, aunque en mi cerebro tengan lugar ciertas actividades que constituyen el hecho de que haya decidido votar a los demócratas, somos más reticentes a atribuir al cerebro este tipo de procesos de decisión racional que, por ejemplo, capacidades perceptivas. No tengo ningún problema en decir que «mi sistema visual sabe distinguir el rojo del violeta», pero soy mucho más reticente a decir que «mi sistema talamocortical ha decidido votar a los demócratas». En el caso de la frase 4, «Soy propietario de un inmueble en la ciudad de Berkeley», no parece que haya nada que podamos abstraer y atribuirlo a la anatomía. Sólo en virtud de mi situación social y las relaciones en que me encuentro puedo ser propietario de un inmueble. El propietario de un inmueble es, en efecto, un cerebro integrado en un cuerpo, pero tal cerebro sólo puede ser propietario de un inmueble en virtud de sus características sociales y legales, y por lo tanto no hay posibilidad de hacer ninguna atribución anatómica o psicológica que sea constitutiva de los aspectos relevantes de los hechos.

Se trata de interesantes cuestiones filosóficas, pero creo que los neurobiólogos las pueden eludir perfectamente. En vez de preocuparse por la medida en que deberían tratar la capacidad de acción racional como una característica de la anatomía neural, deberían concentrarse en el segundo punto, a saber, en el hecho comprobado de que los procesos psicológicos, que constituyen la capacidad de acción racional consciente, tienen lugar en el cerebro y se pueden investigar como tales. Para los fines de la investigación neurobiológica, basta con el estudio del cerebro como mecanismo causal y ubicación anatómica.

He tratado algunos de estos temas filosóficos en otros escritos, como parte del problema del yo.[3] ¿Por qué necesitamos postular un

yo como algo añadido a la secuencia de nuestras experiencias y su realización anatómica? No porque exista ninguna superanatomía más allá de la anatomía ni superexperiencias más allá de la experiencia. No hay más que el cerebro integrado en un cuerpo y las experiencias que tienen lugar en ese cerebro. No obstante, como he argumentado, sí necesitamos postular un yo, pero es una postulación puramente formal. No es un ente adicional sino una especie de principio de organización del cerebro y sus experiencias.

La naturaleza de la filosofía

En el segundo apéndice del libro, Bennett y Hacker critican mis ideas, tanto mi explicación de los problemas de la filosofía de la mente como mi enfoque general de la filosofía. Piensan que mi concepción de la filosofía es errónea en varios aspectos. En mi experiencia, las disputas sobre la naturaleza de la filosofía tienden a ser infructuosas, y normalmente no hacen sino expresar preferencias por diferentes proyectos de investigación. Me eduqué filosóficamente en el Oxford de la década de 1950, donde fui estudiante y profesor de la universidad, y la ortodoxia imperante entonces era que la filosofía trataba del lenguaje y del uso de las palabras. Si alguien dice que la filosofía trata por entero del lenguaje, me parece que con ello expresa una preferencia. Más o menos es equivalente a «Prefiero el trabajo filosófico que se ocupa del lenguaje y me propongo elaborar únicamente trabajo filosófico sobre el lenguaje». Lo que descubrí es que las técnicas que había empleado para analizar el lenguaje funcionaban con otros fenómenos, concretamente los fenómenos mentales y la ontología social. De modo que los métodos que uso son una continuación de los métodos de la filosofía analítica, pero se extienden más allá del do-

minio original de la filosofía del lenguaje y la filosofía lingüística en las que me eduqué.

Un desacuerdo esencial entre mi enfoque y el de Bennett y Hacker es que ellos insisten en que la filosofía no puede ser teórica, no ofrece explicaciones teóricas generales. Todos coincidimos en que la filosofía es, en cierto sentido, esencialmente conceptual, pero la pregunta es: ¿cuál es la naturaleza del análisis conceptual y cuál es su resultado final? Por mi propia experiencia puedo decir que un resultado conceptual es importante sólo como parte de una teoría general. Así que, si observo mi propia biografía intelectual, he propuesto una teoría general de los actos de habla y del significado, una teoría general de la intencionalidad, una teoría general de la racionalidad y una teoría general de la naturaleza de la sociedad, de la ontología social. Si alguien dijera: «Bueno, en filosofía no se pueden tener teorías», mi respuesta sería: «Pues mira, ahí van algunas teorías generales». Un análisis general del acto de prometer, por ejemplo, adquiere mucha más potencia cuando está incorporado a una teoría general del lenguaje y de los actos de habla. Los análisis filosóficos de la acción y la percepción se hacen mucho más profundos cuando se integran en una teoría general de la intencionalidad, y lo mismo ocurre en los demás casos.

En cierto sentido, Bennett y Hacker enuncian erróneamente mis ideas sobre la relación entre la filosofía y la ciencia. Lo que afirmo no es que *todos* los problemas filosóficos puedan convertirse en científicos mediante un análisis conceptual cuidadoso. Al contrario, creo que sólo un reducido número de problemas filosóficos admiten una solución en las ciencias naturales. El problema de la vida es uno de ellos, y confío en que el problema de la relación mente-cuerpo lo sea en el futuro. Pero la mayoría de los problemas que preocupaban a los grandes filósofos griegos, como, por ejemplo, la naturaleza de

la vida buena, la forma de una sociedad justa o el mejor sistema de organización social, no son en mi opinión el tipo de cosas que pueden llegar a ser objeto de tratamiento por parte de las ciencias naturales en ningún modo obvio. Por lo tanto, la suposición de Bennett y Hacker de que a mi entender todos los problemas filosóficos pueden en última instancia convertirse en problemas científicos con soluciones científicas denota una incomprensión de mis tesis por su parte. Tales casos son, desde mi punto de vista, excepcionales.

De nuevo, no creo que sea posible hacer una distinción realmente clara y precisa, como la que ellos afirman haber hecho, entre las cuestiones empíricas y las cuestiones conceptuales y, en consecuencia, no hago una distinción nítida entre cuestiones científicas y cuestiones filosóficas. Permítaseme poner un ejemplo para explicar de qué forma los descubrimientos científicos pueden ayudar a mi trabajo filosófico. Cuando levanto el brazo, mi intención-en-acción* consciente causa un movimiento físico de mi cuerpo. Pero el movimiento también tiene otro nivel de descripción, porque es causado por una secuencia de descargas neuronales y la secreción de acetilcolina en las placas terminales de los axones de las motoneuronas. Partiendo de estos hechos, puedo hacer un análisis filosófico para demostrar que

* La intención-en-acción (*Intention-in-action*) es un concepto introducido por Searle en su análisis de la acción en *Intentionality* (1983). Según Searle, toda acción está compuesta de una intención, que constituye la causa mental, y un movimiento. En las acciones premeditadas, hay una intención previa que causa el acto. La intención-en-acción es el componente intencional de una acción no premeditada que, en contraste con la intención previa, causa el acto estableciendo sobre la marcha, por así decirlo, las condiciones de satisfacción que un movimiento apropiado debe cumplir. Véase Searle, J., *Intencionality. An Essay in the Philosophy of Mind*, Cambridge, Cambridge University Press, 1983 (trad. cast.: *Intencionalidad. Un ensayo en la filosofía de la mente*, Madrid, Tecnos, 1992). *(N. del t.)*

uno y el *mismo* acontecimiento debe ser a la vez un acontecimiento consciente cualitativo y subjetivo, y poseer también muchas propiedades químicas y eléctricas. Pero ahí termina el análisis filosófico. Ahora necesito conocer el mecanismo exacto que hace que esto funcione, exactamente de qué forma el cerebro causa y realiza la intención-en-acción, de tal forma que la intención, con su combinación de estructuras fenomenológicas y electroquímicas, puede mover objetos físicos. Para ello necesitaré los resultados de la investigación neurobiológica.

Bennett y Hacker han escrito un libro importante y, en muchos sentidos, útil. Han invertido mucho trabajo en él. No es mi intención empañar los méritos de su obra con mis objeciones. Sin embargo, creo que la visión que presentan de la neurobiología y de la mente es profundamente errónea y potencialmente perjudicial. Su enfoque proscribiría muchas de las preguntas cruciales que debemos formular en filosofía y en neurociencia. Por ejemplo, ¿qué son los CNC (correlatos neuronales de la conciencia) y cómo causan exactamente la conciencia? ¿Cómo puede mover mi cuerpo mi intención-en-acción consciente? Si se aceptaran las propuestas de Bennett y Hacker, una cantidad enorme de preguntas clave de la investigación neurobiológica serían rechazadas, por considerarlas carentes de significado o incoherentes. Por ejemplo, la pregunta central en el estudio neurocientífico de la visión, «¿De qué forma los procesos neurobiológicos, que se inician con el asalto de los fotones sobre las células fotorreceptoras, y continúan a través del corteza visual hasta adentrase en los lóbulos prefrontales, causan experiencias visuales conscientes?», no la podría investigar nadie que aceptara la concepción de nuestros autores. Se trata de uno de esos casos, como el de

la IA fuerte,* en los que una teoría filosófica equivocada puede tener consecuencias científicas potencialmente desastrosas, y ésta es la razón de que considere importante contestar las tesis de Bennett y Hacker.

* La IA (Inteligencia Artificial) fuerte es una tesis filosófica según la cual lo que llamamos mentes son en realidad programas informáticos digitales. Se habla de IA «fuerte» para distinguirla de la IA «débil» o «cauta», que sostiene que los ordenadores son herramientas muy útiles en el estudio de la mente, pero no que un ordenador tenga mente en sentido literal. Algunos de los corolarios de la IA fuerte son la irrelevancia de la neurofisiología para el estudio de la mente, y la identificación de estados mentales y estados computacionales. Contra esta última tesis dirige Searle su famoso argumento de la Habitación China. Véase Searle, J., «Minds, Brains, and Programs» (1980), en *Behavioral and Brain Sciences*, 3 (3), págs. 417-424 (trad. cast.: «Mentes, cerebros y programas», cap. 22 de Dennett, D. y Hofstader, D. (comps.), *El ojo de la mente: fantasías y reflexiones sobre el yo y el alma*, Buenos Aires, Sudamericana, 1983). *(N. del t.)*

RÉPLICA A LAS REFUTACIONES

LOS SUPUESTOS CONCEPTUALES DE LA NEUROCIENCIA COGNITIVA

Réplica a los críticos

MAXWELL BENNETT Y PETER HACKER

ACLARACIÓN CONCEPTUAL

Nuestro objetivo en *Philosophical Foundations of Neuroscience*[1] era contribuir a la investigación neurocientífica de la única forma en que la filosofía puede asistir a la ciencia: no ofreciendo a los científicos teorías empíricas que reemplacen las suyas, sino esclareciendo las estructuras conceptuales que manejan. Uno de nosotros ha dedicado su vida a construir teorías empíricas sobre las funciones neuronales. Pero esta actividad, en la que se abordan los fundamentos de la neurociencia, no proporciona ninguno de sus fundamentos *conceptuales*. Las aclaraciones sistemáticas que hemos expuesto acerca de la sensación, la percepción, el conocimiento, la memoria, el pensamiento, la imaginación, la emoción, la conciencia y la autoconciencia no son teorías.[2] Su finalidad es esclarecer los conceptos psicológicos que los neurocientíficos cognitivos usan en *sus* teorías empíricas. Las aclaraciones conceptuales que hemos presentado demuestran la existencia de numerosas incoherencias en la actual teorización neurocientífica. Muestran por qué se cometen esos errores y cómo evitarlos.

La neurociencia cognitiva es una investigación experimental que pretende descubrir verdades empíricas referentes a las bases neurales

de las facultades humanas y los procesos neurales que acompañan el ejercicio de tales facultades. Una condición previa de la verdad es el sentido. Si una forma de palabras no tiene sentido, no puede expresar una verdad, y si no expresa una verdad, no puede explicar nada. El objetivo de la investigación filosófica sobre los fundamentos conceptuales de la neurociencia es desvelar y esclarecer verdades conceptuales, verdades que las descripciones plausibles de los descubrimientos y las teorías neurocientíficas presuponen y sin las cuales no tienen sentido.[3] Si esa investigación se lleva a cabo correctamente, esclarecerá los experimentos neurocientíficos y su descripción, además de las inferencias que puedan hacerse a partir de ellos. En *Philosophical Foundations of Neuroscience* delineamos la red conceptual formada por familias de conceptos psicológicos. Estos conceptos están presupuestos por la investigación de la neurociencia cognitiva sobre la base neural de los poderes cognitivos, cogitativos, afectivos y volitivos humanos. Si no se respetan las relaciones lógicas de implicación, exclusión, compatibilidad y presuposición que caracterizan el uso de estos conceptos, lo más probable es que se hagan inferencias no válidas, que se pasen por alto las válidas y que combinaciones de palabras sin sentido sean tratadas como si lo tuvieran.

Algunos filósofos, especialmente en Estados Unidos, han sido muy influidos por la filosofía de la lógica y el lenguaje de Quine, según la cual no hay diferencia significativa entre las verdades empíricas y las conceptuales.[4] Así, desde un punto de vista teórico, el quineano sostendría que no existe una diferencia esencial entre, por ejemplo, la frase «la memoria es conocimiento retenido» y la frase «la memoria depende del funcionamiento normal del hipocampo y la neocorteza». Pero no es así: la primera frase expresa una verdad conceptual; la segunda, un descubrimiento científico. Según Quine, los enunciados de una teoría se confrontan con la experiencia en su

totalidad, y se confirman de forma holística. Pero es un error suponer que la frase «las zorras son hembras» es confirmada por el éxito de la teoría zoológica, o la de «los solteros no están casados» por la sociología de los usos maritales. Asimismo, la afirmación de que el rojo es más oscuro que el rosa, o que el rojo se parece más al naranja que al amarillo, no es verificada por la confirmación de la teoría del color, sino que es un presupuesto de esta teoría. Sería un error suponer que los teoremas del cálculo diferencial fueron confirmados holísticamente por el éxito predictivo de la mecánica newtoniana y hubieran sido debilitados por el fracaso y el rechazo de esta teoría. En realidad, fueron confirmados por *pruebas* matemáticas. Los descubrimientos o las teorías empíricas no pueden confirmar ni debilitar las proposiciones no empíricas, sean proposiciones de la lógica, de las matemáticas o verdades conceptuales manifiestas.[5] Las verdades conceptuales perfilan el espacio lógico en el que se sitúan los hechos. Determinan qué tiene sentido. En consecuencia, los hechos ni las pueden confirmar ni estar en conflicto con ellas.[6]

Una proposición conceptual atribuye propiedades o relaciones internas, una proposición empírica atribuye propiedades o relaciones externas. Una verdad conceptual es en parte constitutiva de los significados de sus expresiones constituyentes, una proposición empírica es una descripción de cómo son las cosas. Una verdad conceptual es una afirmación implícita de una norma descriptiva disfrazada de afirmación fáctica. Precisamente porque este tipo de afirmaciones son en parte constitutivas de los significados de sus expresiones constituyentes, la incapacidad de reconocer una verdad conceptual (por ejemplo, que el rojo es más oscuro que el rosa) es un criterio indicativo de la falta de comprensión de una u otra de sus expresiones constituyentes.

Es evidente que esta concepción normativa de la verdad conceptual no se refiere *especialmente* a las llamadas proposiciones ana-

líticas en ninguna de las diversas, familiares y *diferentes* concepciones de la analiticidad, como las de Kant, Bolzano, Frege y Carnap. De hecho, las diversas distinciones entre lo analítico y lo sintético simplemente se soslayan. En su lugar, distinguimos entre la afirmación de una medida y la afirmación de una medición. Suponer que la distinción que establecemos entre lo conceptual y lo empírico es una distinción *epistémica* sería una muestra de total incomprensión.[7] Esta distinción *no* se infiere de *cómo conocemos* las respectivas verdades, sino del *papel* de la proposición en cuestión: si es normativa (y constitutiva) o descriptiva. Hay que destacar que el hecho de que sea una cosa o la otra es una característica del uso de un enunciado, no (o no necesariamente) de un tipo de enunciado. El mismo enunciado usado en un contexto para expresar una verdad conceptual puede usarse en otro contexto como una afirmación de hecho —como bien se demuestra en la mecánica newtoniana—. En muchos contextos, es posible que no esté claro, *a falta de una investigación ulterior*, qué papel se supone que desempeña una frase en uso. De hecho, en la ciencia, lo típico es que las pruebas inductivas y las constitutivas (los criterios lógicos) fluctúen. Pero está muy claro que caracterizar una frase como expresión de una verdad conceptual es destacar su *función* distintiva como afirmación de una medida, más que de una medición. De ahí que la distinción, a diferencia de la distinción *a priori/a posteriori*, no sea epistemológica, sino lógica o lógico-gramatical.

Dos paradigmas: Aristóteles y Descartes

La reflexión filosófica sobre la naturaleza humana, sobre el cuerpo y el alma, se remonta a los albores de la filosofía. Platón y Aristóteles fijaron las polaridades entre las que esa reflexión fluctúa. Según Pla-

tón, y la tradición platónico-cristiana de san Agustín, el ser humano no es una sustancia unificada, sino una combinación de dos sustancias distintas, un cuerpo mortal y un alma inmortal. Según Aristóteles, el ser humano *es* una sustancia unificada, en la que el alma (*psuchē*) es la forma del cuerpo. Describir esta forma es describir las capacidades características de los seres humanos, en particular las capacidades distintivas del intelecto y la voluntad que caracterizan a la *psuchē* racional. El debate moderno sobre este tema se inicia con la heredera de la tradición platónica-agustiniana, a saber, la concepción cartesiana de los seres humanos como dos cosas en una, una mente y un cuerpo. Su interacción causal recíproca se invocó para explicar la experiencia y la conducta humanas.

Las figuras de mayor importancia de las primeras dos generaciones de neurocientíficos del siglo XX, por ejemplo, Sherrington, Eccles y Penfiel, eran dualistas cartesianos confesos. La tercera generación conservó la estructura cartesiana básica, pero la transformó en dualismo cuerpo-cerebro: se abandonó el dualismo sustancial pero se mantuvo el dualismo estructural. Los neurocientíficos contemporáneos atribuyen al cerebro el mismo conjunto de predicados mentales que Descartes atribuía a la mente, y conciben en gran medida la relación entre el pensamiento y la acción, y la experiencia y sus objetos, del mismo modo en que lo hacía Descartes exceptuando, esencialmente, la mera sustitución de la mente por el cerebro. El tema central de nuestro libro era demostrar la incoherencia del dualismo cerebro/cuerpo y revelar su erróneo carácter criptocartesiano. La parte constructiva de nuestro propósito era demostrar que, para hacer justicia a la estructura de nuestro esquema conceptual y ofrecer descripciones coherentes de los grandes descubrimientos de la neurociencia cognitiva post-sherringtoniana, es necesaria una explicación aristotélica, con el debido énfasis en las capacidades

activas y pasivas de primer y segundo orden y sus manifestaciones conductuales, complementada con intuiciones de Wittgenstein que se añaden a las de Aristóteles.[8]

El principio de Aristóteles y la falacia mereológica

En *Philosophical Foundations of Neuroscience* señalamos un error persistente que llamamos «la falacia mereológica de la neurociencia».[9] Corregir este error es un leitmotiv (pero *sólo* un leitmotiv) de nuestro libro. Decimos que el error es «mereológico» porque implica adscribir a las partes atributos que sólo se pueden adscribir de modo inteligible a las totalidades de las que son partes. Aristóteles señaló una forma de este error en torno al año 350 a.C., cuando observaba que «afirmar [...] que es el alma [*psuchē*] quien se irrita, sería como afirmar que es el alma la que teje o edifica. Mejor sería, en realidad, no decir que es el alma quien se compadece, aprende o discurre, sino el hombre en virtud del alma» (DA 408^{b} 12-15)* —hacer algo en virtud de la propia alma es lo mismo que hacerlo en virtud de las propias capacidades—. Es un error adscribir al alma de un animal atributos que sólo se pueden adscribir literalmente al animal en su conjunto. Podemos llamar a este principio «el principio de Aristóteles».

Nuestro interés primordial estaba en el equivalente neurocientífico de ese principio, a saber, el error de atribuir al *cerebro* —una parte de un animal— atributos que literalmente sólo se pueden adscribir al animal en su totalidad. No fuimos los primeros en señalarlo; Anthony Kenny lo había hecho en su brillante artículo de 1971

* Aristóteles, *Acerca del alma*, Madrid, Gredos, 1978, pág. 155, trad. de Tomás Calvo Martínez. *(N. del t.)*

«The Homunculus Fallacy».[10] Este error es más propiamente *mereológico* que su ancestro aristotélico, ya que el cerebro es literalmente una parte del animal sensitivo mientras que, en contra de las afirmaciones de Platón y Descartes, el alma o mente no lo es. Hoy observamos, con espíritu aristotélico, que decir que el cerebro está irritado es como decir que el cerebro teje o edifica. Y es que sin duda es mejor no decir que el cerebro se compadece, aprende o discurre, sino que una persona hace todas estas cosas.[11] En consecuencia, negamos que tenga sentido decir que el cerebro es consciente, tiene sensaciones, percibe, piensa, sabe o quiere cualquier cosa, ya que éstos son atributos de los animales, no de sus cerebros.

Nos sorprendió un poco que el profesor Dennett crea que lo que *nosotros* teníamos en mente era la distinción, expuesta en su *Content and Conscious* de 1969,* entre los niveles personal y subpersonal de las explicaciones. Allí afirmaba, correctamente, que padecer un dolor no es una propiedad del cerebro. Pero la razón que daba es que los dolores son «fenómenos mentales» que son «no mecánicos», mientras que los procesos cerebrales son «esencialmente mecánicos».[12] El contraste que *nosotros* establecimos entre las propiedades del todo y las de las partes no es el que hay entre lo no mecánico y lo mecánico. Lo que marca las horas es el reloj en su totalidad, no sus ejes ni sus engranajes —aunque el proceso de marcar las horas es completamente mecánico—. Lo que vuela es el avión, no sus motores —aunque el proceso de volar es completamente mecánico—. Además, los verbos de sensación, como «doler», «picar» o «cosquillear», *sí* se aplican a las partes de un animal, cuya pierna puede dolerle, su cabeza picarle, o puede que tenga cosquillas en las ijadas (PFN, pág. 73). Estos atributos son, en palabras del profesor Dennett, «no mecánicos»;

* Trad. cast.: *Contenido y conciencia*, Barcelona, Gedisa, 1996. *(N. del t.)*

no obstante *son* atribuibles a partes del animal. De modo que nuestra observación mereológica es muy distinta de la distinción que el profesor Dennett establece entre los niveles personal y subpersonal de la explicación, y, aplicada a los animales, muy diferente de su distinción entre lo que es «mecánico» y lo que no lo es.[13]

¿La falacia mereológica es realmente mereológica?

El profesor Searle objetaba que lo que hemos caracterizado como un paradigma de una falacia mereológica, es decir, la adscripción de atributos psicológicos al cerebro, no es tal, ya que el cerebro no es una parte de la persona, sino más bien una parte del cuerpo de una persona (pág. 107). Esto, a nuestro entender, no es más que marear la perdiz. Las palabras de Wittgenstein que citábamos eran: «Sólo del *ser humano* y de lo que se parece a un ser humano (se comporta como tal) se puede decir: tiene sensaciones; ve, es ciego; oye, es sordo; es consciente o inconsciente»[14] (la cursiva es nuestra). El cerebro *es* una parte del ser humano.

El profesor Searle indica que, si adscribir atributos psicológicos al cerebro realmente fuera un error mereológico, tal error se desvanecería si esos atributos se adscribieran a lo que él llama «el resto del sistema» al que pertenece el cerebro. Cree que el «resto del sistema» es el cuerpo que *tiene* el ser humano. Observa que no atribuimos predicados psicológicos al cuerpo que tenemos. Con la notabilísima excepción de los verbos que indican sensación (por ejemplo, «Me duele todo el cuerpo»), el último punto es correcto. No decimos: «Mi cuerpo percibe, piensa o sabe». Sin embargo, «el sistema» al que se puede decir que el cerebro humano pertenece como una parte es *el ser humano*. El cerebro humano es una parte del ser humano, del mismo modo

que el cerebro canino es una parte del perro. Mi cerebro —el cerebro que tengo— es una parte de mí —del ser humano viviente que soy— tanto como lo son mis piernas y mis brazos. Pero es verdad que también se puede decir que mi cerebro es una parte de mi cuerpo.

¿Cómo se explica esto? Cuando hablamos de nuestra mente lo hacemos en gran medida en un sentido *no agentivo* y *metafórico*, refiriéndonos a nuestras capacidades racionales del intelecto y la voluntad, y a su ejercicio. Cuando hablamos del cuerpo, hablamos de nuestras propiedades corporales. Hablar de *mi cuerpo* es hablar de los rasgos corporales *del ser humano que soy* —unos rasgos que pertenecen al aspecto (un cuerpo atractivo o desgarbado)—, a las superficies del ser humano (su cuerpo estaba cubierto de picaduras de mosquito, todo él lacerado, pintado de azul), a aspectos de la salud y el estado físico (un cuerpo sano o enfermo) y, sorprendentemente, a la sensación (me puede doler todo el cuerpo, como me duele la pierna o me pica la espalda).[15] Pero saber, percibir, pensar, imaginar, etc., no son características corporales de los seres humanos, y no son atribuibles al cuerpo que el ser humano tiene, del mismo modo que no son atribuibles al cerebro que el ser humano tiene. Los seres humanos no son *sus cuerpos*. No obstante, *son* cuerpos, en el sentido completamente distinto de ser una clase particular de persistente espaciotemporal sensitivo —el *homo sapiens*— , y el cerebro es una parte del ser humano viviente, como lo son las extremidades.[16] Sin embargo, no es una parte consciente, pensante y perceptiva, como no lo es cualquier otra parte del ser humano. Y es que éstos son atributos del ser humano como un todo.

No obstante, el profesor Searle ha señalado un rasgo interesante de nuestro lenguaje metafórico referido al cuerpo. Los seres humanos son personas —es decir, animales inteligentes y capaces de usar lenguaje—, son autoconscientes, poseen conocimiento del bien y

del mal, son responsables de sus actos y tienen derechos y obligaciones. Ser persona, por decirlo someramente, es poseer esas capacidades que acreditan el estatus de agente moral. Probablemente no diríamos que el cerebro es parte de la persona sino que es parte del cuerpo de la persona, mientras que no dudaríamos en decir que el cerebro de Pepe es una parte de Pepe, parte de *este* ser humano, del mismo modo, exactamente, que las piernas y los brazos de Pepe son partes de Pepe. ¿Por qué? Tal vez porque «persona» es, como subrayó Locke, «un término forense», pero no un nombre-sustancia. Por tanto, si usamos el término «persona» en contextos como éste, con ello indicamos que lo que nos interesa en primer lugar son los seres humanos en cuanto poseedores de aquellas características que los hacen personas, con relativa independencia de las características corporales. Quizás ayude la siguiente analogía: Londres es una parte del Reino Unido; el Reino Unido pertenece a la Unión Europea, pero Londres no. Esto no impide que Londres forme parte del Reino Unido. Así también el hecho de que Pepe sea una persona no impide que su cerebro sea parte de él.

El fundamento del principio

¿Por qué deberíamos aceptar el principio de Aristóteles y su equivalente neurocientífico? ¿Por qué deberíamos desalentar a los neurocientíficos que se empeñan en atribuir *conciencia*, *conocimiento*, *percepción*, etc. al cerebro?

La *conciencia*. Son los animales quienes son conscientes o inconscientes, y quienes pueden llegar a ser conscientes de algo que capte su atención. Es el estudiante, no su cerebro, quien se despierta y se hace consciente de lo que está diciendo el profesor, y es el

profesor, no su cerebro, quien es consciente del aburrimiento de sus alumnos al verlos bostezar con disimulo. El cerebro no es el órgano de la conciencia. Uno ve con sus ojos y oye con sus oídos, pero no es consciente con su cerebro, del mismo modo que no anda con él.

Un animal puede ser consciente sin demostrarlo. Éste es el *único* sentido en que se puede decir, con el profesor Searle, que «la propia existencia de la conciencia no tiene nada que ver con la conducta» (pág. 104). Pero *el concepto* de conciencia está estrechamente ligado a las evidencias conductuales para atribuir conciencia al animal. Un animal no tiene que exhibir cierta conducta para *ser* consciente. Pero sólo de un animal al que *se pueda atribuir en un modo inteligible* esa conducta se puede afirmar también que es consciente, y que esta afirmación sea *verdadera o falsa*. No tiene sentido atribuir conciencia o razonamiento a una silla o una ostra, porque no hay silla ni ostra que se queden dormidas para después despertarse, o que pierdan la conciencia y la recuperen de nuevo; y no hay silla ni ostra que se comporten de manera razonable o poco razonable.[17] La «cuestión ontológica» (como la llama el profesor Searle) —la cuestión de la verdad (como nosotros preferimos llamarla)— presupone la previa determinación de la cuestión del sentido. El acuerdo sobre las evidencias conductuales para la atribución de la conciencia, es decir, sobre lo que *cuenta* como una manifestación de conciencia, es una precondición para la investigación científica sobre los requisitos neurales para ser consciente. De no ser así, no se podría ni siquiera identificar qué se quiere investigar. Distinguir la cuestión del sentido de la cuestión de la verdad no significa confundir «las reglas para el uso de las palabras con la ontología», como señala el profesor Searle (pág. 105), al contrario, significa diferenciarlas.[18]

El profesor Searle insiste en que la conciencia es una propiedad del cerebro. Sherrington, Eccles y Penfield, al ser cartesianos, pensaban

erróneamente que es una propiedad de la mente. ¿Qué experimento neurocientífico reciente puede aportar el profesor Searle para demostrar que *realmente* es una propiedad del cerebro? Al fin y al cabo, lo único que los neurocientíficos *podrían* descubrir es que ciertos estados neurales están correlacionados inductivamente con el hecho de que un *animal* sea consciente, y son condiciones causales de ello. Pero *este* descubrimiento no puede demostrar que es *el cerebro* lo que es consciente. ¿Ocurre, entonces, que la afirmación del profesor Searle es una reflexión conceptual? No —porque no es éste el modo en que se utiliza el concepto de *ser consciente*—. Son los seres humanos (y otros animales), no sus cerebros (ni sus mentes), los que se quedan dormidos para despertarse después, los que pueden ser noqueados y después recuperar la conciencia. Por lo tanto, ¿es una recomendación lingüística, concretamente que, cuando el cerebro de un ser humano está en un estado inductivamente correlacionado con el hecho de que el ser humano es consciente, deberíamos describir su cerebro también como consciente? Es una convención que podríamos adoptar. Podríamos introducir este uso derivativo de «ser consciente». Sería necesariamente dependiente del uso primario que se aplica al ser humano en su totalidad. Sin embargo, es difícil encontrar ningún motivo que lo haga aconsejable. Desde luego no aporta nada a la claridad de la descripción, y no añade nada más que una forma vacía a la explicación neurocientífica existente.

El *conocimiento*. El conocimiento comprende capacidades de diversos tipos. La identidad de una capacidad está determinada por aquello de que es capacidad. La razón más simple para atribuir una capacidad a un animal es el hecho de que lleva a cabo actividades corporales que manifiestan esa capacidad. Cuanto más compleja es la capacidad, más diversas y difusas son las razones para atribuirla. Si un animal sabe algo, puede actuar y reaccionar ante su entorno en mo-

dos de los que sería incapaz si lo ignorara; si exhibe estos modos en su conducta, demuestra su conocimiento. Se puede decir que el cerebro es el *vehículo* de esas capacidades, pero lo que esto significa es que, en ausencia de las estructuras neurales apropiadas, el animal no sería capaz de hacer lo que hace. Las estructuras neurales del cerebro son diferentes de las capacidades que tiene *el animal*, y el funcionamiento de estas estructuras es distinto del ejercicio de las capacidades *por parte del animal*. En resumen, el que sabe es también el que actúa, y su saber se muestra en su actuar.

Criticamos a J. Z. Young por sostener, como hacen muchos neurocientíficos, que el cerebro contiene conocimiento e información «del mismo modo que el conocimiento y la información están registrados en libros u ordenadores».[19] El profesor Dennett asegura que no hemos hecho nada para establecer que no exista un concepto del conocimiento o de la información del que no se pueda decir que esté codificado tanto en los libros como en el cerebro (pág. 91). De hecho sí que examinamos el asunto (PFN, págs. 152 y sigs.). Pero lo explicaremos de nuevo.

Un código es un sistema de convenciones de cifrado y transmisión de información dependiente del lenguaje. Un código no es un lenguaje, no tiene gramática ni léxico (por ej., el código Morse). El conocimiento no está *codificado* en los libros, a no ser que estén escritos en código. Sólo se puede codificar un mensaje si existe un código en el que hacerlo, y sólo existe un código si los codificadores y los descodificadores a los que se dirigen se ponen de acuerdo sobre las convenciones de la codificación. En este sentido, no existe, y no podría existir, un código neural. El cerebro no contiene información en el sentido en que la contiene un libro. El cerebro no posee información en el sentido en que la posee un ser humano. El hecho de que de rasgos del cerebro se pueda derivar información (como

del tronco de un árbol se puede derivar información dendrocronológica) no demuestra que la información esté codificada en el cerebro (del mismo modo que no lo está en el tronco del árbol).

Por consiguiente, en el sentido usual de «conocimiento», el cerebro no puede registrar, contener ni poseer conocimiento. En este punto, el profesor Dennett cambia de táctica, y recomienda que atendamos a la literatura científica cognitiva sobre las *extensiones* del término «conocimiento» que podrían permitir que el conocimiento, en un sentido ampliado, se atribuyera al cerebro. Y recomienda que nos fijemos en el intento de Chomsky de explicar un concepto extendido del conocimiento, a saber el de «cognoscimiento», según el cual los seres humanos, incluso los neonatos, *cognoscen* los principios de la gramática universal.[20] Según Chomsky, alguien que *cognosce* no puede decirle a otro lo que *cognosce*, no puede mostrar el objeto de su *cognoscimiento*, no reconoce lo que *cognosce* cuando se le describe, nunca olvida lo que *cognosce* (pero tampoco lo recuerda nunca), nunca lo ha aprendido, y no podría enseñarlo. Aparte de esto, *cognoscer* es lo mismo que conocer. ¿Se postula *esto* como modelo de extensión inteligible de un término?

La *percepción*. Las facultades perceptivas son capacidades de adquirir conocimientos mediante el uso de los propios sistemas sensoriales. El animal usa los ojos para echar un vistazo, contemplar, atisbar y mirar cosas. Así es capaz de distinguir las cosas que tienen colores, forma y movimientos distintivos. Demuestra su agudeza visual con lo que hace como respuesta a lo que ve. No poseería estas capacidades perceptivas ni sería capaz de ejercerlas de no ser por el funcionamiento adecuado de las partes apropiadas de su cerebro. Sin embargo, no es la corteza cerebral la que ve, sino el animal. No es el cerebro el que se aproxima a las cosas para verlas mejor, el que mira entre los arbustos y bajo los setos. No es el cerebro el que salta

para huir de un depredador que ha detectado, ni el que ataca la presa que ha visto, es el animal dotado de percepción quien hace todo esto. En pocas palabras, el que percibe es también el que actúa.

En *Consciousness Explained*, el profesor Dennett adscribe atributos psicológicos al cerebro. Afirma que éste es consciente, reúne información, hace suposiciones simples, se basa en la información de que dispone y llega a conclusiones.[21]

Esto significa caer en la misma falacia de la que advierten tanto Aristóteles como Wittgenstein: la falacia mereológica, tal como la hemos llamado. En su artículo para la APA, el profesor Dennett admite que sería una falacia atribuir predicados psicológicos *cabales* a partes del cerebro (pág. 87). No obstante, sostiene, es teóricamente útil y coherente con la aceptación del carácter erróneo de atribuir predicados del todo a sus partes, extender el vocabulario psicológico, *debidamente atenuado*, de los seres humanos y otros animales a) a los ordenadores y b) a partes del cerebro. En efecto, sostiene abiertamente que no existe ninguna diferencia significativa entre estas dos extensiones. Pero hay una diferencia: atribuir propiedades (atenuadas o no) a los ordenadores es un error, pero no implica una falacia mereológica. Considerar que el cerebro es como un ordenador y atribuirle, a todo él o a sus partes, tales propiedades psicológicas es doblemente erróneo. Vamos a explicarlo.

Es verdad que, en el habla informal, decimos que los ordenadores recuerdan, que buscan en su memoria, que calculan y, a veces, cuando tardan mucho, decimos en broma que están pensando. Pero esto no es más que un modo de hablar, no una aplicación literal de los términos «recordar», «calcular» y «pensar». Los ordenadores son aparatos diseñados para desempeñar determinadas funciones en nuestro beneficio. En un ordenador podemos almacenar información, como podemos hacerlo en un archivo. Pero los archivos no pueden

recordar nada, como tampoco pueden los ordenadores. Utilizamos ordenadores para obtener el resultado de un cálculo —del mismo modo que antes usábamos una regla de cálculo o calculadora mecánica—. Estos resultados se obtienen sin que nadie ni nada literalmente calcule — como resulta obvio en el caso de una regla de cálculo o calculadora mecánica—. Para calcular literalmente hay que dominar una amplia variedad de conceptos, seguir una multitud de reglas que uno debe conocer y comprender toda una diversidad de operaciones. Los ordenadores no lo hacen ni lo pueden hacer.

El profesor Dennett sugiere que «es un hecho empírico [...] que *partes* de nuestro cerebro llevan a cabo procesos *asombrosamente* parecidos al conjeturar, decidir, creer, llegar a conclusiones, etc. Y se parecen *lo suficiente* a estas conductas en el nivel personal para justificar la extensión del lenguaje corriente para abarcarlos» (pág. 86). Conviene en que sería un error «atribuir creencias», decisiones, deseos o sufrimientos *cabales* al cerebro. Mejor es decir que «del mismo modo que un niño pequeño puede *algo así como* creer que su papá es médico [...] cierta parte del cerebro de una persona puede *algo así como* creer que hay una puerta abierta unos metros más adelante» (pág. 87).

Esto es parte de lo que el profesor Dennett caracteriza como la «actitud intencional» —una metodología de investigación que supuestamente ayuda a los neurocientíficos a explicar los fundamentos neurales de las capacidades humanas—. Dice Dennett que la adopción de la actitud intencional ha permitido realizar «un excelente trabajo científico [...] generar hipótesis verificables, desarrollar teorías, analizar fenómenos enormemente complejos en sus partes más comprensibles» (pág. 87). Parece que acepta la idea de que algunas partes del cerebro «algo así como creen», que otras *algo así como deciden*, y aún otras *algo así como supervisan* estas actividades. Se supone, presumiblemente, que todo esto *algo así como explica* lo que los neuro-

científicos quieren explicar. Pero si lo explicado son una especie de creencias, pseudoexpectativas, protodeseos y semidecisiones (como Dennett sugiere [pág. 88]), en el mejor de los casos sólo alcanzan a tener *algo así como sentido* y cabe presumir que sólo son *algo así como verdad*. Y cómo se pueden hacer inferencias válidas a partir de estas premisas es más que algo así como oscuro. Exactamente cómo se supone que tales premisas *explican* los fenómenos es igualmente oscuro. Y es que la lógica de estas presuntas explicaciones es totalmente confusa. ¿Es que las algo así como creencias, las pseudocreencias, las protocreencias o las semicreencias sobre algo proporcionan a una parte del cerebro una razón para actuar? ¿O sólo algo así como una razón? —¿para algo así como una acción?—. Cuando a Dennett se le pregunta si las partes del cerebro son, como él dice, «auténticos sistemas intencionales», su respuesta es: «No preguntéis» (pág. 88).[22]

Los neurocientíficos cognitivos formulan *auténticas* preguntas preguntan *cómo* está implicada la corteza prefrontal en el pensamiento humano, *por qué* existen vías de reentrada, *cuáles* son exactamente los papeles respectivos del hipocampo y la neocorteza en la memoria humana. Decir que el hipocampo algo así como recuerda durante breves momentos, y que la neocorteza tiene en cambio algo así como memoria a largo plazo, no es ningún tipo de explicación. De las explicaciones de Dennett no ha surgido ninguna teoría empírica confirmada en la neurociencia, porque atribuir «algo así como propiedades psicológicas» a partes del cerebro no *explica* nada. Volveremos a este tema cuando hablemos de la explicación que Sperry y Gazzaniga dan de la comisurotomía. No sólo no aporta explicación alguna, sino que genera más incoherencia.[23]

Coincidimos con el profesor Dennett en que muchas de las creencias del niño lo son en un sentido atenuado. La comprensión que una niña pequeña tiene del concepto de un médico puede ser de-

fectuosa, pero la niña dirá acertadamente: «Mi papá es médico», y a la pregunta: «¿Dónde está el médico» responderá: «Allí dentro» (señalando el despacho de su papá). De manera que se puede decir que la niña cree, en un sentido atenuado, que su padre es médico. En su conducta verbal y deíctica, la niña cumple *algunos* de los criterios normales de la creencia que su padre es médico (pero también cumple algunos de los criterios de ausencia de tal creencia). Pero no existe nada parecido a una parte de un cerebro que afirma cosas (como hace la niña), que responde preguntas (como hace la niña), ni que señala cosas deícticamente (como hace la niña). Así pues, en el sentido en que la niña, en su conducta verbal y deíctica, puede manifestar una creencia *rudimentaria*, una parte del cerebro no puede hacerlo, del mismo modo que el cerebro en su totalidad no puede manifestar una creencia cabal. ¿O es que el profesor Dennett puede sugerir un experimento crucial que demuestre que la corteza prefrontal de la niña algo así como cree que el gato está debajo del sofá?

La niña también puede mostrar una creencia rudimentaria en su conducta no verbal. Si ve que el gato corre a esconderse debajo del sofá y ella, con su andar aún inseguro, se acerca a buscarlo ahí, se puede decir que cree que el gato está debajo del sofá. Pero el cerebro y sus partes no pueden *comportarse* de ningún modo, no pueden acercarse con andar inseguro al sofá, no pueden mirar debajo de éste, y no pueden parecer desconcertados si bajo el sofá no hay gato alguno. Las partes del cerebro no pueden actuar voluntariamente ni emprender acciones en vistas a un fin. A diferencia de la niña, las partes del cerebro no pueden cumplir *ninguno* de los criterios para creer algo, ni siquiera en un sentido rudimentario. El cerebro (y sus partes) sólo pueden «algo así como creer» en el mismo sentido en que son «algo así como océanos» (ya que existen ondas cerebrales) y «algo así como sistemas climáticos» (ya que existen las tormentas de

ideas). Entre los procesos del cerebro y las creencias, decisiones o conjeturas de los seres humanos hay tanta similitud como la que hay entre el cerebro y el mar. Después de todo, tanto el cerebro como el mar son de color gris, tienen arrugas en su superficie y por ellos circulan corrientes.

La ubicación de los atributos psicológicos

La pregunta de si el cerebro es un posible sujeto de atributos psicológicos es distinta de la de si es el lugar de aquellos atributos psicológicos a los que se les puede asignar inteligiblemente una ubicación corporal (PFN, págs. 122 y sigs. y 179 y sigs.) Las razones que nos llevaban a negar que el cerebro pueda ser el sujeto de atributos psicológicos no demuestran que el cerebro no sea el lugar de aquellos atributos a los que tiene sentido asignar una ubicación corporal. Tampoco pretendían hacerlo. Desde nuestra perspectiva *es posible* asignar una ubicación a sensaciones como el dolor y el picor. La ubicación de un dolor está donde señala quien lo sufre, en el miembro que se masajea, en la parte del cuerpo que dice que le duele —porque son estas formas de la conducta del dolor las que aportan los criterios para la ubicación de éste—. Por el contrario, al pensar, creer, decidir y desear, por ejemplo, no se les puede asignar una ubicación *somática*. La respuesta a las preguntas «¿Dónde pensaste en esto?», «¿Dónde adquirió esa extraña creencia?», «¿Dónde tomó la decisión de casarse?» *nunca* es «En la corteza prefrontal, por supuesto». Los criterios para determinar dónde un ser humano pensó en algo, adquirió una creencia, tomó una decisión, se enfadó o se quedó estupefacto implican una conducta, sin duda, pero no una conducta indicativa de ubicación somática. La ubicación del pensar, recordar, ver, decidir,

enfadarse o asombrarse por parte de un ser humano está *donde el ser humano esté cuando piensa,* etc.[24] Qué parte de su cerebro interviene en su hacer esas cosas es otra importante cuestión que los neurocientíficos poco a poco van entendiendo mejor. Pero no están descubriendo dónde tiene lugar el pensar, el recordar o el decidir, están descubriendo qué partes de la corteza están causalmente implicadas en el pensar, el recordar o el decidir del ser humano.

Por supuesto, pensar en algo, decidir hacer algo, ver algo son, como acertadamente dice el profesor Searle (pág. 110), acontecimientos reales, realmente ocurren en algún lugar del mundo y en algún momento. Se me ocurrió este argumento en la biblioteca y decidí cómo formularlo en mi despacho, vi a Pepe cuando estaba en la calle y escuché el recital de Inés en el auditorio. El profesor Searle sugiere que la pregunta «¿Dónde ocurren los acontecimientos mentales?» no es más filosóficamente abstrusa que la pregunta «¿Dónde se producen los procesos digestivos?». Los procesos digestivos, argumenta, ocurren en el estómago y la conciencia ocurre en el cerebro. Es un error. Estar consciente, en oposición a estar inconsciente, ser consciente de algo, en oposición a no percatarse de ello o no prestarle atención, no se producen *en* el cerebro en modo alguno. Ocurren, por supuesto, *debido a ciertos acontecimientos en el cerebro*, sin los cuales el ser humano no habría recuperado la conciencia ni su atención se habría visto atraída. Para responder la pregunta «¿Dónde te hiciste consciente del sonido del reloj?» tengo que especificar dónde estaba cuando tal sonido me llamó la atención, del mismo modo que para responder la pregunta «¿Dónde recuperaste la conciencia?» tengo que especificar dónde estaba cuando volví en mí.

Tanto el digerir como el pensar se predican de los animales, pero de ello no se sigue que no haya diferencias lógicas entre ambos. Se puede decir que el estómago digiere los alimentos, pero no que el

cerebro esté pensando. El estómago es el órgano digestivo, pero el cerebro no es un órgano del pensamiento al igual que no lo es de la locomoción.[25] Si se abre el estómago, se puede ver cómo tiene lugar la digestión en él. Pero si se quiere ver cómo tiene lugar el pensar, habría que observar *El pensador* de Rodin (o al cirujano durante la operación, el jugador de ajedrez durante la partida o el contertulio en el calor del debate), no al cerebro. Todo lo que el cerebro de una persona puede mostrar es qué pasa ahí *mientras la persona está pensando*; todo lo que las imágenes por RMf pueden mostrar es qué parte del cerebro de esa persona está metabolizando más oxígeno que las demás cuando el paciente en cuestión está pensando.[26] (Atribuimos a las vigas de hierro longitud, fuerza y presencia de grietas. Pero de ahí no se sigue que la *longitud* y la *fuerza* tengan el mismo carácter lógico; y podemos decir dónde está la grieta, pero no dónde está la fuerza.)

Así pues, las sensaciones, por ejemplo los dolores, *están* ubicadas en nuestro cuerpo. Pero el profesor Searle sostiene que todas ellas están *en el cerebro*. Él mismo admite que es algo que va contra la intuición —al fin y al cabo, nos quejamos de dolor de estómago, de gota en los pies o de artritis en las rodillas— . No obstante, afirma, el cerebro crea una imagen corporal, y el dolor del que decimos que está en el pie, y que para aliviarlo nos masajeamos el pie, es una conciencia-del-dolor-como-localizado-en-el-pie, que se encuentra en la imagen del cuerpo que hay en nuestro cerebro. Es interesante que Descartes adoptara una concepción muy parecida, al señalar que «el dolor de la mano lo siente el alma no porque esté presente en la mano, sino porque está presente en el cerebro».[27] La ventaja de su explicación, señala el profesor Searle, es que significa que podemos describir el fenómeno del dolor fantasma sin caer en el absurdo de sugerir que el dolor está en el espacio físico, en la cama o debajo de la

sábana. Caemos en ese mismo absurdo, sostiene, cuando decimos que los dolores están en el cuerpo. Estamos de acuerdo en que es un absurdo, pero no en que caigamos en él.

Hay muchos usos locativos de «en», algunos espaciales, otros no espaciales («en el cuento», «en octubre», «en comité»). Entre los usos espaciales, hay muchos tipos diferentes, en función de qué está en qué (PFN, págs. 123 y sigs.). Estamos de acuerdo con el profesor Searle en que si hay una moneda en el bolsillo de mi chaqueta y mi chaqueta está en el vestidor, entonces hay una moneda en el vestidor. Pero no todos los usos locativos de «en» son transitivos en este sentido. Si hay un agujero en mi chaqueta y la chaqueta está en el armario, no se sigue que haya un agujero en el armario. En el caso de la chaqueta y la moneda, *nos ocupamos* de relaciones espaciales entre dos objetos independientes, pero no ocurre lo mismo en el caso de la chaqueta y el agujero. Asimismo, si hay una arruga en mi camisa, y mi camisa está en la maleta, no se sigue que haya una arruga en la maleta. La moneda se puede sacar del bolsillo de la chaqueta, y la camisa se puede sacar de la maleta, pero el agujero no se puede sacar del bolsillo, hay que zurcirlo, como hay que planchar la arruga, no sacarla.

El uso de «en» respecto a la ubicación de las sensaciones no es como con la moneda, sino más bien como con el agujero (aunque, de todos modos, diferente). Un dolor no es una sustancia. Si siento un dolor en el pie, *no me sitúo en ninguna relación* con un dolor —al contrario, mi pie me duele *ahí*, y puedo señalar el lugar que duele, al que llamamos «la ubicación del dolor»—. En el caso del miembro fantasma, a quien sufre el dolor le parece que el miembro amputado sigue en su lugar, y reporta un dolor en el miembro ilusorio. Le parece como si le doliera la pierna, pese a no tener pierna. Coincidimos con el profesor Searle en que no es la cama la que due-

le y que el dolor que el amputado siente no está debajo de la sábana. El hecho de que éste sienta el dolor donde estaría su pierna y que su pierna estaría debajo de la sábana, no implica que haya un dolor debajo de la sábana, al igual que el hecho de que tenga dolor en su pierna no amputada y ésta esté en su bota no implica que tenga dolor en la bota. En realidad, convenimos con el profesor Searle en lo que se refiere a los fenómenos, y sólo discrepamos en lo referente a su descripción. El hecho de que el dolor del amputado sea real pero que la ubicación en que lo siente sea ilusoria (no le duele la pierna, pues no tiene pierna) no demuestra que cuando una persona que *no* haya sufrido una amputación sienta un dolor en la pierna la ubicación en que lo siente sea también ilusoria. Realmente es su pierna lo que le duele. No creemos que haya en el cerebro unas imágenes del cuerpo y nos preguntamos qué pruebas hay de su existencia —al fin y al cabo, si se abre el cerebro de un ser humano no se encuentra en él ninguna imagen del cuerpo—. A lo que al parecer se refiere el profesor Searle es a que los métodos fisiológicos, empezando por los de Sherrington, se han utilizado para establecer que las neuronas de la corteza somatosensorial se pueden excitar en una relación de uno-a-uno con puntos estimulados de la superficie del cuerpo, y con la distribución espacial de los músculos de las extremidades y el tronco. Pero es totalmente confuso qué quiere decir el profesor Searle cuando habla de «tener un dolor en un pie fantasma fenomenológico en la imagen del cuerpo que está en el cerebro» (pág. 118 y sigs.) Uno *puede* tener dolor en la *cabeza*, comúnmente conocido como dolor de cabeza. Pero no puede tener dolor de espalda o de estómago en el cerebro, ni cualquier otro tipo de dolor. Y no es ninguna casualidad, ya que allí no hay terminaciones nerviosas salvo en la duramadre.

Por último, el profesor Searle sostiene que cuando un filósofo dice que dos personas sienten el mismo dolor, lo que quiere decir es

que tienen el mismo tipo de dolor, pero son dos ejemplares de dolor distintos (pág. 116). «[C]omo ejemplar, el dolor que ellos o yo experimentamos *existe sólo en tanto que lo percibe* un sujeto consciente en particular» (pág. 116). Falso. En primer lugar, los dolores no los *perciben* quienes los sufren. *Tener* un dolor no es *percibir* un dolor. «Tengo un dolor en la pierna» no describe una relación entre yo y un objeto llamado «un dolor» más de lo que lo hace la frase equivalente «Me duele la pierna». En segundo lugar, la distinción de Peirce entre tipo y ejemplar se aplicaba a las inscripciones y depende de las convenciones ortográficas. No es de aplicación a los dolores más de lo que pueda serlo a los colores. Si dos sillones son de color granate, entonces hay dos sillones exactamente del mismo color, y no dos ejemplares del mismo tipo. Porque ¿cómo se pueden individualizar los diferentes ejemplares? Obviamente no por la ubicación —ya que ésta meramente distingue los dos sillones, no su color—. Todo lo que se puede decir es que el primer supuesto ejemplar pertenece al primer sillón, y el segundo, al segundo sillón. Pero esto significa individualizar una propiedad por referencia a la pseudopropiedad de pertenecer a la sustancia que la tiene —como si las propiedades fueran sustancias que se distinguen mediante la ley de Leibniz, y como si ser la propiedad de una determinada sustancia fuera una propiedad que distingue, por ejemplo, el color de *esta* silla del color de *aquélla*—. Y esto es absurdo. Las dos sillas son exactamente del mismo color. Asimismo, si dos personas tienen un terrible dolor de cabeza en la sien izquierda, entonces ambas tienen exactamente el mismo dolor. El dolor de A no se distingue del dolor de B por el hecho de que pertenece a A, del mismo modo que el color granate del primer sillón no se distingue del color granate del segundo sillón por el hecho de que pertenece al primero. La distinción entre la identidad cualitativa y la numérica no es de aplicación a los colores ni a los do-

lores, como tampoco lo es la distinción de Peirce entre tipos y ejemplares.

ANTROPOLOGÍA LINGÜÍSTICA, AUTOANTROPOLOGÍA, METÁFORA Y EXTENSIÓN DEL USO

El profesor Dennett señala que examinar el uso de las palabras precisa bien una forma de antropología, bien una «autoantropología». En el primer caso hay que descubrir los usos de las palabras mediante adecuadas encuestas sociológicas, pidiendo a las personas que se fijen en sus intuiciones sobre el uso correcto de una palabra. Otra posibilidad es consultar las propias intuiciones; pero podría resultar que las intuiciones de uno sean divergentes de las de los demás. El profesor Dennett afirma que nosotros no consultamos a la comunidad de neurocientíficos para averiguar las intuiciones de éstos sobre la jerga neurocientífica de los predicados psicológicos (pág. 86 y nota al pie 15, pág. 204), sino sólo nuestras propias intuiciones.

Esto es un error. Un hablante competente de la lengua no tiene que consultar sus intuiciones (presentimientos, conjeturas), al igual que un matemático competente no tiene que consultar sus intuiciones sobre las tablas de multiplicar, ni un jugador de ajedrez competente debe consultar sus intuiciones sobre el movimiento de las piezas. Es un hecho empírico, que los antropólogos, los historiadores de la lengua, etc. deben establecer, que una determinada palabra hablada o escrita se use o se usase de una forma determinada en una determinada comunidad lingüística. No es un hecho empírico que una palabra, que *tiene el significado que tiene*, tenga las conexiones conceptuales, las compatibilidades y las incompatibilidades que tiene. Es

un hecho empírico que el vocablo «negro» lo utilizan los hispanohablantes para decir lo que quiere decir, pero en tanto que significa lo que significa, a saber, *este* ☞ ■ color, no es un hecho empírico que las proposiciones «El negro es más oscuro que el blanco», «El negro se parece más al gris que al blanco», «Nada puede ser totalmente negro y totalmente blanco a la vez» sean verdaderas. Son verdades conceptuales, que especifican una parte de la red conceptual de la que *negro* es un nodo. La incapacidad de reconocer estas verdades denota que no se ha captado en absoluto el significado de la palabra. El hablante competente es aquel que ha llegado a dominar el uso de las expresiones comunes de la lengua. No son intuiciones suyas que negro es *ese* ☞ ■ color, que yegua es la hembra del caballo o que deambular es pasear. No es un presentimiento suyo que un hombre es un ser humano adulto de sexo masculino. Y no son conjeturas suyas que si son las diez es más tarde de las nueve, ni que si algo es totalmente negro no es totalmente blanco.

Aunque los hablantes competentes de una lengua coinciden en la lengua que hablan, las desviaciones del uso común no son, *como tales*, filosóficamente nocivas. Tales desviaciones pueden no ser más que un fragmento de un idiolecto personal o un sociolecto especial, una extensión nueva de un término o la apropiación de un término existente para un nuevo uso técnico. Esto es lo que nos llevó a decir que si un hablante competente emplea expresiones contrarias al uso común, entonces bien pudiera ocurrir que

> no haya que entender sus palabras en su significado habitual. Tal vez las expresiones problemáticas se usaron en un sentido especial, y en realidad son meramente homónimas; o eran extensiones *analógicas* del uso acostumbrado, tan comunes en la ciencia; o se usaron en un sentido metafórico o figurativo. Si se puede recurrir a estas vías de escape,

> la acusación de que los neurocientíficos han sido víctimas de la falacia mereológica está injustificada.
>
> (PFN, pág. 74)

Pero que estas vías de escape estén disponibles no es algo que se pueda dar por supuesto. Tampoco la conveniencia de esta aplicación del término es un asunto sobre el que el hablante en cuestión sea la última autoridad. Porque, aun en el caso de que esté introduciendo un uso nuevo, o de que esté empleando sus palabras en sentido figurativo, hay que ver si lo hace con coherencia. Y hay que investigar si lo hace de forma consistente o si, por el contrario, se mueve inadvertidamente entre un uso nuevo y otro viejo, haciendo inferencias a partir del primero que sólo se justificarían desde el segundo. Por esta razón decíamos que

> La autoridad última sobre el asunto es *el propio razonamiento del hablante*. Debemos fijarnos en las consecuencias que saca de sus propias palabras, y serán las inferencias que haga las que demostrarán si estaba usando el predicado en un nuevo sentido o si hacía de él un mal uso. Si hay que condenarle, tiene que ser por su propia boca.
>
> (PFN, pág. 74)

Y, a continuación, pasábamos a demostrar que a muchos neurocientíficos importantes sin duda se les podría condenar por su propia boca, precisamente porque hacen inferencias a partir de su aplicación peculiar del vocabulario psicológico al cerebro, unas inferencias que sólo se pueden hacer de forma inteligible a partir de la aplicación habitual de ese vocabulario al animal en su totalidad (PFN, caps. 3-8).

Si un neurocientífico aplica al cerebro expresiones psicológicas o semánticas, como «representaciones» o «mapa», o bien está usando

estas expresiones en su sentido habitual o no lo está haciendo. Si no lo está haciendo, entonces ese uso no habitual puede implicar 1) un sentido derivativo, 2) una extensión analógica o de otro tipo del término antiguo, 3) un simple homónimo o 4) un sentido metafórico o figurativo. Si los términos psicológicos se aplican al cerebro en su sentido habitual, lo que se dice no es inteligible. No sabemos qué significa decir que el cerebro piensa, teme o se avergüenza. Los criterios constitutivos de la aplicación, por parte de los hablantes competentes de nuestra lengua, de tales expresiones a los animales y los seres humanos, a saber, lo que dicen y hacen, no los pueden cumplir el cerebro ni sus partes —no existe nada parecido a un cerebro o parte de él que haga observaciones prudentes, que huya por miedo o se sonroje de vergüenza—. No comprendemos qué supondría que los cerebros o sus partes pensaran, temieran o decidieran algo, mejor de lo que comprendemos qué supondría que lo hicieran los árboles. Si tales términos se aplican en un sentido nuevo, entonces el usuario nos debe una explicación de cuál es ese sentido. Puede ser un sentido derivativo, como cuando aplicamos el término «sano» a los alimentos o al ejercicio físico —un uso que requiere una explicación distinta de la adecuada explicación del uso primario de «salud» aplicado a un ser vivo—. Puede ser un uso analógico, como cuando hablamos del pie de la montaña o de la página —tales analogías suelen ser evidentes, pero está claro que exigen una explicación parafrástica muy distinta de la que requiere su prototipo—. O puede ser un homónimo, como «masa» en la mecánica de Newton, que requiere una explicación muy distinta de la de «masa» en «toda una masa de gente» o en «con las manos en la masa».

Cuando los neurocientíficos hablan de «representación», se refieren, en su mayor parte, a un simple homónimo de «representación» en su sentido simbólico y semántico. Este uso ha resultado in-

fausto, pues ha llevado a eminentes científicos y psicólogos a sucumbir a la confusión de utilizar la palabra tanto para referirse a un correlato causal o concomitante como para referirse a una representación simbólica. Pero sólo en el primer sentido se puede hablar de representaciones en el cerebro. De ahí nuestra crítica a David Marr (PFN, págs. 70, 76, 143-147). Parece que el uso que los neurocientíficos hacen del término «mapa» [*map*] inició su andadura como una extensión de la idea de «mapeo» [*mapping*], pero enseguida se confundió con la idea de un mapa. Nada hay de erróneo en decir que un conjunto de entes sobre el que se pueden mapear miembros de otro conjunto es un «mapa» del último —aunque no es un uso necesario ni resulta del todo claro—. Pero la incoherencia surge cuando se supone que el cerebro podría utilizar este «mapa» a la manera en que los lectores de un atlas usan los mapas. No está claro en absoluto qué se quiere decir cuando se afirma que el cerebro o sus partes saben, creen, piensan, infieren y perciben cosas. La única idea coherente que podría apuntarse con ello es que estos términos se aplican al cerebro para significar la actividad neural que supuestamente se corresponde con el saber, el creer, el pensar, el inferir y el percibir del animal. Pero de ahí no se sigue en ningún sentido inteligible la afirmación (que hacen Crick, Sperry y Gazzaniga) de que la parte del cerebro que está pensando comunica *lo que piensa* a otra parte del cerebro. Porque, si el bien el pensamiento humano tiene un contenido (que proporciona la respuesta a la pregunta «¿Qué estás pensando?»), no se puede decir que la actividad neural tenga ningún tipo de contenido.

Se podría sugerir que cuando los neurocientíficos hablan de *mapas* o *descripciones simbólicas* en el cerebro, y de que éste sabe, piensa, decide, interpreta, etc., lo hacen de forma metafórica.[28] Podría afirmarse que estos términos son en realidad agudas metáforas cuya

idoneidad ha sido establecida hace ya mucho en lo que se refiere a los *ordenadores electrónicos*, de los que se dice certeramente que «siguen reglas». Los ordenadores «*se construyeron deliberadamente* para llevar a cabo la "manipulación de símbolos complejos según reglas"». En realidad, podría pensarse que esa forma de hablar «ya no es ni siquiera metafórica, considerando el marco teórico y tecnológico altamente desarrollado en que tiene lugar».[29] Asimismo, en el uso que hacen del vocabulario psicológico común, los neurocientíficos cognitivos «están de hecho andando a tientas en la oscuridad; las metáforas son la regla, más que la excepción». Pero esto es parte del progreso científico normal, y en su avance la neurociencia ha ido más allá de la metáfora en algunos casos, por ejemplo en la adscripción al cerebro de «representaciones tipo frase» y «representaciones tipo mapa».

Todo esto se puede cuestionar. No es correcto decir que los ordenadores siguen reglas, al igual que no se puede decir que los planetas obedecen leyes. Las leyes de Kepler describen el movimiento de las órbitas de los planetas, pero éstos no *cumplen* las leyes. Los ordenadores no se construyeron «para llevar a cabo la manipulación de símbolos según reglas», se construyeron para producir unos resultados que *coincidan* con la *correcta* manipulación de símbolos según reglas. Los ordenadores no pueden *seguir* una regla más de lo que pueda hacerlo una calculadora mecánica. Una máquina puede realizar operaciones que se ajusten a una regla, siempre y cuando todas las conexiones causales integradas en ella funcionen como se diseñó que lo hicieran, y suponiendo que el diseño asegura la generación de una regularidad acorde con la regla o reglas escogidas. Pero para que algo constituya un caso de seguir una regla, no basta la mera producción de una regularidad acorde con una regla. Se puede decir que un ser sigue una regla sólo en el contexto de una práctica compleja que implica las actividades reales y potenciales de justificar los errores, per-

catarse de ellos y corregirlos tomando como referencia la regla, criticar las desviaciones de la regla y, si es requerido a ello, explicar que una acción cumple la regla y enseñar a otros en qué consiste seguir una regla. La determinación de que un acto es *correcto*, que está de acuerdo con la regla, no es una determinación causal, sino lógica. De lo contrario, deberíamos rendirnos ante cualesquiera que fueran los resultados que los ordenadores producen.[30]

Indudablemente, los ingenieros informáticos utilizan este tipo de lenguaje —al principio de forma inocua, hasta que llega el momento en que empiezan a aplicarlo literalmente y suponen que los ordenadores realmente piensan, que lo hacen mejor y más deprisa que nosotros, que realmente recuerdan y a diferencia de nosotros nunca olvidan, y que interpretan lo que tecleamos y a veces lo interpretan mal, asignando a lo que escribimos un significado distinto del que pretendíamos expresar—. *Entonces* la manera de hablar, por lo demás inocua, de los informáticos deja de ser un modo divertido de abreviar y se convierte en una confusión conceptual potencialmente perniciosa.

Es *posible* que se quiera decir que los ordenadores o el cerebro piensan, calculan, razonan, infieren y formulan hipótesis sólo metafóricamente. Las metáforas no explican —ilustran una cosa en términos de otra—. Una metáfora deja de serlo cuando se convierte en un cliché, como «un corazón roto» o «de un tirón», pero no puede dejar de ser una metáfora convirtiéndose en literal. ¿Qué significaría que es *literalmente* verdadero que los planetas obedecen las leyes de la naturaleza? Se puede decir que una regla de cálculo, una calculadora mecánica o un ordenador calculan —hablando figuradamente—. ¿Pero que significaría que literalmente calcularan? Si «el ordenador recuerda (calcula, infiere, etc.)» tiene perfecto sentido (no figurativo) para los ingenieros informáticos es precisamente porque emplean estas frases como clichés. «El ordenador calcula» no signi-

fica nada más que «El ordenador pasa por el proceso electromecánico necesario para producir los resultados de un cálculo sin cálculo alguno», del mismo modo que «Te quiero con todo mi corazón» no significa nada más que «Te quiero de verdad».

Es notorio que cuando los neurocientíficos hablan, en los casos que hemos criticado, de que en el cerebro hay unos mapas, y de que el cerebro usa estos mapas como tales, no lo hacen en sentido metafórico. Resulta obvio que la observación de Blakemore de que «los neuroanatomistas dicen ahora que el cerebro tiene *mapas*, que se cree desempeñan un papel fundamental en la representación y la interpretación que el cerebro hace del mundo, *del mismo modo que los mapas de un atlas lo hacen para sus lectores*»[31](la cursiva es nuestra) no es metafórica, ya que nada hay de metafórico en «Los mapas del atlas desempeñan un papel en la representación del mundo de sus lectores». Además, el término «representación» se emplea aquí manifiestamente en el sentido simbólico, no en el sentido de correlato causal. Tampoco es metafórica la afirmación de J. Z. Young de que el cerebro hace uso de sus mapas al formular sus hipótesis sobre lo que es visible.[32] Porque *hacer uso de un mapa al formular una hipótesis* significa tomar una característica indicada por el mapa como *razón* para la hipótesis. El profesor Dennett afirma que el cerebro «*hace* uso de ellos *como* mapas» (pág. 205, n. 20), y en el debate de la APA aseguró que es una cuestión empírica, y no filosófica, la de si el cerebro usa los «mapas retinotópicos» *como mapas*, la de «si alguna de las operaciones de recuperación de información que se definen al hablar de ellos explota las características de los mapas que nosotros explotamos cuando lo hacemos con los mapas del mundo real». Pero sólo podría ser una cuestión empírica si tuviera sentido que el cerebro usa un mapa como un mapa. Sin embargo, para *usar* un mapa como un mapa, tiene que *haber* un mapa —y en el cerebro no hay

ninguno—; uno ha de saber leer el mapa—pero el cerebro no tiene ojos y no puede leer—; hay que estar familiarizado con las convenciones de proyección del mapa (por ejemplo, proyección cilíndrica, cónica, de azimut) —pero no existen convenciones *de proyección* referentes al mapeo de las características del campo visual sobre las activaciones neurales en la corteza estriada «visual»; y uno debe usar el mapa para guiarse en su conducta —en sus excursiones o navegaciones—, actividad que el cerebro no realiza. No hay que confundir un mapa con la posibilidad de un mapeo. El hecho de que se pueda mapear la activación de las células retinales sobre la activación de las células de la corteza estriada visual no demuestra que haya un mapa de lo visible en el campo visual de la corteza estriada.

Por último, quisiéramos rectificar un malentendido. Algunos de nuestros críticos presumen que intentamos imponer una ley que prohíba las extensiones nuevas de las expresiones de la lengua. El profesor Dennett afirmó en el debate de la APA que proscribiríamos el uso de «*código* genético», puesto que insistimos en que no se puede *codificar* el lenguaje en el cerebro. El profesor Churchland supone que excluiríamos por principio innovaciones conceptuales como la introducida por Newton cuando decía que la Luna *cae* constantemente hacia la Tierra a medida que se mueve sobre su trayecto inercial. Todo esto es un malentendido.

No prohibimos nada —sólo apuntamos las incoherencias conceptuales que se producen en los escritos neurocientíficos—. No tratamos de impedir a nadie que extienda los usos del lenguaje de forma científicamente provechosa —únicamente tratamos de asegurarnos de que tales supuestas extensiones no traspasan los límites del sentido por una incapacidad para especificar el uso nuevo, o por mezclarlo indebidamente con el antiguo—. Nada hay de malo en hablar del *pie* de la montaña, siempre y cuando uno no se pregunte

si la montaña lleva zapatos. Nada hay de malo en hablar del *paso* del tiempo, siempre y cuando uno no se confunda (como se sabe que le ocurrió a san Agustín) sobre cómo medirlo. No había nada de malo en que Newton dijera que la Luna «cae», pero lo habría habido de haberse preguntado qué era lo que la hacía resbalar. Nada tenía de malo que hablara de fuerzas que actúan sobre un cuerpo en el espacio, pero lo habría habido si hubiese especulado sobre si las fuerzas eran de infantería o de caballería. Nada tiene de malo que los genetistas hablen del código genético. Pero lo habría si de la existencia de éste hicieran inferencias que sólo se pueden hacer de la existencia de códigos literales. Y es que, en realidad, el código genético no es un código en el sentido en que se usa un código para cifrar o transmitir una frase de una lengua. No es ni siquiera un código en «un sentido atenuado», como se podría decir de una frase acordada entre los padres para que los hijos no sepan de qué están hablando.

Estábamos preocupados por el uso que los neurocientíficos cognitivos hacen del vocabulario psicológico común y corriente (y de otros términos, como «representación» y «mapa») al especificar aquellos fenómenos que sus teorías pretenden explicar y describir los términos con los que se proponen hacerlo. Porque, como hemos dejado claro, habitualmente los neurocientíficos intentan explicar el percibir, el saber, el creer, el recordar y el decidir de los seres humanos haciendo referencia a partes del cerebro que perciben, saben, creen, recuerdan y deciden. Por esto citábamos observaciones, obra de importantes neurocientíficos, psicólogos y científicos cognitivos, como las siguientes:

> J. Z. Young: «Podemos considerar toda visión como una búsqueda continua de respuestas a preguntas que formula el cerebro. Las señales de la retina constituyen "mensajes" que contienen esas respues-

tas. Luego, el cerebro usa esta información para construir una hipótesis adecuada de lo que hay ahí».

Programs of the Brain, pág. 119

C. Blakemore: «El cerebro [tiene] mapas, que se cree desempeñan un papel esencial en la representación y la interpretación que el cerebro hace del mundo, del mismo modo que los mapas de un atlas lo hacen para sus lectores».

«Understanding Images in the Brain», pág. 265

G. Edelman: «El cerebro "relaciona de forma recursiva las secuencias semánticas con las fonológicas, y luego genera unas correspondencias sintácticas [...] tratando para ello a las reglas que se desarrollan en la memoria como objetos de manipulación conceptual"».

Bright Air, Brilliant Fire (Harmondsworth, Penguin, 1994), pág. 130

J. Frisby: «En el cerebro debe de haber una descripción simbólica del mundo exterior, una descripción formulada en símbolos que se refieren a los diversos aspectos del mundo de los que la vista nos hace conscientes».

Seeing: Illusion, Brain, and Mind (Oxford, Oxford University Press, 1980), pág. 8

F. Crick: «Cuando se corta el cuerpo calloso, el hemisferio izquierdo sólo ve la mitad derecha del campo visual [...] ambos hemisferios pueden oír lo que se dice [...] parece que una mitad del cerebro ignora por completo lo que la otra mitad vio».

The Astonishing Hypothesis (Londres, Touchstone, 1995), pág. 170

S. Zeki: «La capacidad del cerebro de adquirir conocimientos, de abstraer y de construir ideales».

«Splendours and Miseries of the Brain»,
en *Philosophical Transactions of the Royal Society B* 354 (1999),
pág. 2.054

D. Marr: «Nuestros cerebros de algún modo han de ser capaces de representar [...] información [...] Por consiguiente, el estudio de la visión debe incluir [...] una investigación sobre la naturaleza de las representaciones internas por las que captamos esta información y la hacemos accesible como base para las decisiones [...]

Una representación es un esquema formal para describir [...] junto con las reglas que especifican cómo hay que aplicar el esquema [...] [Un esquema formal es] un conjunto de símbolos con reglas para combinarlos [...] Una representación, por lo tanto, no es en absoluto una idea extraña: todos usamos representaciones continuamente».

Vision (San Francisco, Freeman, 1980), págs. 20 y sigs.

Éstos no son usos metafóricos. No son extensiones audaces de los términos que introducen significados nuevos con fines teóricos. Son sencillamente usos incorrectos del vocabulario psicológico (y semántico) común —unos usos incorrectos que conducen a la incoherencia y a diversas formas de contrasentido— que hemos señalado caso por caso. Nada hay de extraño en ello. No difiere, en principio, de las aplicaciones igualmente equivocadas del mismo vocabulario *a la mente,* como si fuera mi mente la que sabe, cree, piensa, percibe, siente dolor, quiere y decide. Pero no es así; soy yo, el ser humano viviente, quien hace todo eso. El primer error no es menos mayúsculo que el último (y venerable) error, y está extendido

entre los neurocientíficos cognitivos, a veces en detrimento de los experimentos que éstos diseñan, habitualmente en detrimento de su teorizar los resultados de esos experimentos y muy a menudo en detrimento de su explicación de las funciones cognitivas animales y humanas por referencia a las estructuras y las operaciones neurales que las hacen posibles.

LOS *QUALIA*

En nuestra exposición de la conciencia (PFN, caps. 9-12), argumentábamos que caracterizar el dominio de lo mental por referencia a la «sensación cualitativa» de la experiencia es erróneo (PFN, cap. 10). Pero, con permiso del profesor Searle (pág. 99 y sigs.), no negamos la existencia de los *qualia* aduciendo que, si existieran, existirían en el cerebro. Si, *per impossibile*, todos los atributos psicológicos se caracterizaran por su «sensación cualitativa», seguirían siendo atributos de los seres humanos, no del cerebro.

Se supone que un *quale* es «la sensación cualitativa de una experiencia» (Chalmers),[33] o es algo como «la rojez del rojo o lo doloroso del dolor» (Crick).[34] Los *qualia* son «las simples cualidades sensoriales que se encuentran en lo azul del cielo o en el sonido de una nota» (Damasio)[35] o «lo que se siente al ver, oír u oler, lo que se siente al tener un dolor» (Block).[36] Según el profesor Searle, los estados conscientes son «cualitativos en el sentido de que, para cada estado consciente [...] hay algo que cualitativamente es como estar en ese estado».[37] Según Nagel, para cada experiencia consciente «hay algo que es como tenerlo para el organismo».[38] Estas diversas explicaciones *no son equivalentes*, y no está claro que a partir de ellas emerja una explicación coherente.

El profesor Searle observa que en un dolor, un picor o unas cosquillas hay una sensación cualitativa. Estamos de acuerdo, en el sentido siguiente: las sensaciones, señalábamos (PFN, pág. 124), poseen cualidades fenoménicas (por ejemplo, las de arder, escocer, picar, roer, punzar); están asociadas a la inclinación a ciertas conductas (soplarse, rascarse, frotarse o reírse); y tienen grados de intensidad que pueden aumentar o disminuir.

Sin embargo, observábamos, cuando se trata de la percepción es difícil caracterizar lo que se quiere decir con «el carácter cualitativo de la experiencia». Especificar lo que vemos u olemos o, en el caso de las alucinaciones, lo que nos parece que vemos u olemos, requiere que se especifique un objeto. Las experiencias visuales u olfativas y sus correspondientes alucinaciones se individualizan por aquello de lo que son experiencias o alucinaciones. Ver una farola es distinto de ver un buzón, oler una lila es distinto de oler una rosa, y así lo son también las correspondientes experiencias alucinatorias, que se describen en términos de la similitud que presentan para el sujeto con su contrapartida perceptiva verídica.[39]

Las rosas, sin duda, no huelen como las lilas —cómo huelen las rosas es distinto de cómo huelen las lilas—. Oler rosas es muy diferente de oler lilas. Pero el carácter cualitativo de oler rosas no huele a rosas, ni el carácter cualitativo de oler lilas huele a lila. El oler unas u otras puede ser igualmente agradable —en cuyo caso el carácter cualitativo del oler puede ser exactamente el mismo, aunque lo que se huela sea distinto—. Creemos que el profesor Searle confunde cómo se sienten los olores con cómo se siente el oler.

Normalmente, al ver una farola no se siente nada. Si se pregunta «¿Qué sintió al verla?» el único tipo de respuesta es del estilo «No sentí nada en particular, ni agradable ni desagradable, ni excitante ni aburrido». *Estos* epítetos —«agradable», «desagradable», «ex-

citante», «aburrido»— se *entienden* correctamente cómo descripciones del «carácter cualitativo de la experiencia». En este sentido, *muchas* experiencias perceptivas no tienen ningún carácter cualitativo. *Ninguna* se individualiza por su sensación cualitativa —se individualizan por su objeto—. Y si hablamos de una alucinación, decir que la farola de nuestra alucinación era de color negro sigue siendo una descripción del objeto de la experiencia —su «objeto intencional» en la jerga de Brentano (que el profesor Searle utiliza)—. Por otro lado, la cualidad de la experiencia alucinatoria es probablemente *sobrecogedora.*

Contrariamente a lo que sugiere el profesor Searle, nosotros no argumentamos que «si los *qualia* no se definen como una cuestión de agradabilidad o repugnancia, entonces habrá que individualizar la experiencia por su objeto» (pág. 115). Nuestra tesis era que *sí* individualizamos las experiencias y las alucinaciones por sus objetos —que se especifican por la respuesta a la pregunta «¿*De* qué fue experiencia (o alucinación) tu experiencia (o alucinación)?»—.[40] Claro está que no es necesario que el objeto sea la causa, como se pone de manifiesto en el caso de las alucinaciones. Pero, insistíamos, no se debe confundir el carácter cualitativo de la experiencia con las cualidades del objeto de la experiencia. El hecho de que lo que uno ve al ver una manzana roja sea rojo y redondo no implica que uno tuviera una experiencia visual roja y redonda. El hecho de que lo que a uno le parece que ve cuando alucina una manzana roja sea rojo y redondo no implica que uno tuviera una alucinación visual roja y redonda. «¿Qué viste (o alucinaste)?» es una pregunta, y «¿Cómo fue ver lo que viste (o alucinar lo que alucinaste?» otra. Las experiencias perceptivas no se individualizan por su carácter cualitativo. Se trata de verdades sencillas, pero al parecer se han pasado por alto.

Cerebros dentro de un cráneo

El profesor Searle dice que los seres humanos son «cerebros dentro de un cuerpo» (págs. 120 y sigs.). Desde su punto de vista, la razón de que podamos decir tanto «Peso 75 kilos» como «Mi cuerpo pesa 75 kilos» es que lo que hace que yo pese 75 kilos es que eso es lo que pesa mi cuerpo. Pero, al parecer, estrictamente hablando no soy más que un cerebro dentro de un cuerpo (de un cráneo). *Tengo* un cuerpo, y estoy *en* el cráneo *de* mi cuerpo. Es ésta una versión materialista del cartesianismo. Una de las razones principales de que escribiéramos nuestro libro fue la firme creencia de que los neurocientíficos contemporáneos, y también muchos filósofos, siguen aún bajo la oscura y alargada sombra de Descartes. Al tiempo que rechazan la sustancia inmaterial de la mente cartesiana, transfieren los atributos de la mente cartesiana al cerebro humano, dejando intacta la totalidad de la estructura de la concepción cartesiana de la relación entre la mente y el cuerpo. Lo que nosotros defendemos es que los neurocientíficos, e incluso los filósofos, dejen atrás las sombrías regiones cartesianas y busquen la luz del sol aristotélico, donde se puede ver mucho mejor.

Si, *per impossibile*, yo fuera un cerebro dentro de un cuerpo, entonces tendría un cuerpo —del mismo modo que la mente cartesiana tiene un cuerpo—. Pero no *tendría* cerebro, ya que los cerebros no tienen cerebros. Y en verdad mi cuerpo no pesaría 75 kilos, sino 75 kilos menos los 1,5 kilos que, hablando con propiedad, es lo que *yo* pesaría. Y no mediría 1,75 de alto, sino sólo 0,18 metros. No cabe duda de que el profesor Searle me asegurará que soy mi-cerebro-dentro-de-mi-cuerpo —mi cerebro *junto* con mi cuerpo—. Pero esto no nos lleva a ninguna parte. Y es que mi cerebro junto con mi cuerpo sin cerebro, tomado en un sentido, es simplemente mi cadáver; tomado en otro sentido, es sencillamente *mi cuerpo*. Pero yo

no soy mi cuerpo, no el cuerpo que *yo tengo*. Naturalmente, soy *un* cuerpo —el ser humano viviente que está ante el lector, un tipo particular de persistente espaciotemporal sensible que posee inteligencia y voluntad y, por consiguiente, es una persona—. Pero no soy *mi cuerpo* más de lo que soy *mi mente*. Y tampoco soy un cerebro dentro de un cuerpo. Es un error suponer que los seres humanos están «dentro de cuerpos» —esta idea pertenece a la tradición platónica, agustiniana y cartesiana que se debería rechazar—. Sería mucho mejor decir, con Aristóteles, que los seres humanos son criaturas *animadas* (*empsuchos*), animales dotados de unas capacidades que les confieren, en la forma de vida que les es natural, el estatus de personas.

La investigación neurocientífica

Nuestros críticos afirman que nuestras investigaciones son irrelevantes para la neurociencia o, peor aún, que nuestro consejo, de seguirse, resultaría netamente perjudicial. El profesor Dennett sostiene que nuestro rechazo a adscribir atributos psicológicos (incluso en un sentido atenuado) a nada que no sea un animal en su totalidad es retrógrado y anticientífico, en contraste con los beneficios científicos de la «actitud intencional» que él defiende. En su opinión, «la licencia poética que permite la actitud intencional facilita la tarea» de explicar cómo el funcionamiento de las partes contribuye a la conducta del animal (pág. 89).

En primer lugar, observamos que la *licencia poética* es algo que se permite a los poetas para los fines de la poesía, no para los fines de la precisión empírica y el poder explicativo. En segundo lugar, atribuir capacidades cognitivas a partes del cerebro sólo ofrece la apariencia

de una explicación cuando se requiere una explicación verdadera y, por tanto, obstaculiza el proceso científico. Sperry y Gazzaniga sostienen que, en los casos de comisurotomía, la extraña conducta de los sujetos experimentales expuestos a representaciones de objetos se *explica* por el hecho de que un hemisferio del cerebro ignora lo que la otra mitad puede ver.* Se presume que los hemisferios del cerebro conocen cosas y las pueden explicar, y que, debido a la sección del cuerpo calloso, el hemisferio derecho no puede comunicar al izquierdo lo que ve, de modo que este último debe generar su propia interpretación de por qué la mano izquierda está haciendo lo que hace.[41] Lejos de explicar los fenómenos, esto no es más que redescribirlos en términos confusos, obviando así la ausencia de una explicación sustancial. La disociación de funciones normalmente asociadas se explica en parte por la sección del cuerpo calloso y por la ubicación de funciones en ambos hemisferios. *Esto* se sabe hoy a la perfección, pero la explicación actualmente disponible no va más allá. Es una ilusión suponer que nada en absoluto se añade a esa explicación con la atribución de («algo así como», u otro tipo)

* En una célebre serie de experimentos, Gazzaniga, Bogen y Sperry (1962) presentaron a pacientes comisurotomizados estímulos visuales y táctiles dirigidos alternativamente a uno de los dos hemisferios «incomunicados» por la ausencia del cuerpo calloso. Los estímulos visuales presentados al campo visual derecho (conectado al hemisferio izquierdo) eran correctamente reportados por los sujetos, pero cuando los mismos estímulos se presentaban al campo visual izquierdo (conectado al hemisferio derecho) los sujetos afirmaban no haber visto nada. Si a continuación se les pedía que cogieran un objeto usando la mano izquierda (controlada por el mismo hemisferio derecho), lo hacían sin problema, pero si se les preguntaba entonces qué tenían en la mano, los sujetos eran incapaces de responder. Véase Gazzaniga, M. S., Bogen, J. E., y Sperry, R. W. (1962) «Some functional effects of sectioning the cerebral commissures in man», en *Proceedings of the National Academy of Sciences*, 48, parte 2, pág. 1.765. *(N. del t.)*

conocimiento, percepción y comprensión lingüística a los hemisferios del cerebro.

El profesor Searle afirma que, si nuestra explicación de las estructuras conceptuales utilizadas fuera correcta, se rechazarían, por carecer de significado, las preguntas esenciales de la investigación neurobiológica. Por esto, apunta que «la pregunta central de la visión, de qué forma los procesos neurobiológicos [...] causan experiencias visuales conscientes, no la podría investigar nadie que aceptara [nuestra] concepción». Nuestra concepción, asegura, «puede tener unas consecuencias científicas potencialmente desastrosas» (pág. 124).

La investigación sobre la neurobiología de la visión es la investigación sobre las estructuras neurales causalmente necesarias para que un animal pueda ver, y sobre los procesos específicos implicados en la visión. Nuestro rechazo a la idea de que las experiencias visuales ocurren en el cerebro, o que se caracterizan por los *qualia*, afecta a este programa de investigación neurocientífica sólo en la medida en que evita preguntas fútiles que podrían no tener respuesta. Damos numerosos ejemplos: el problema de la integración (Crick, Kandel y Wurtz), la explicación del reconocimiento por referencia a la comparación de plantillas e imágenes (Marr) o la sugerencia de que las percepciones son hipótesis del cerebro, conclusiones de las inferencias inconscientes que hace (Helmholtz, Gregory y Blakemore). Nuestra concepción de que quien ve o tiene experiencias visuales es el animal, no el cerebro, y la concepción del profesor Searle de que es el cerebro, no el animal, son tesis conceptuales, no empíricas. El tema es sin duda importante por todo lo dicho, pero debe quedar claro que lo que nosotros decíamos no *dificulta* la investigación empírica sobre los procesos neurales en que se sustenta la visión. Al contrario, dirige la descripción de los resultados de tales investigaciones hacia el camino del sentido.

En general, las críticas conceptuales de nuestro libro no hacen más que arrancar capas de confusión conceptual de los estudios neurocientíficos, y esclarecer las formas conceptuales que éstos presuponen. Esto no puede impedir el progreso de la neurociencia. En realidad, debería facilitarlo, mediante la exclusión de preguntas carentes de sentido, la proscripción de experimentos desacertados y la depuración de los resultados experimentales malinterpretados.[42]

EPÍLOGO

MAXWELL BENNETT

Conocí a sir John Eccles en 1962, cuando estaba terminando mis estudios de licenciatura en ingeniería eléctrica. Al año siguiente obtuvo el premio Nobel por su trabajo sobre la transmisión química en las sinapsis de la médula espinal y el cerebro. Me preguntó qué estaba haciendo, y yo le dije que «ingeniería eléctrica», a lo que replicó: «Excelente, debes venirte conmigo, ya que todo laboratorio neurofisiológico de primera clase necesita un buen soldador». Creo que todo laboratorio de neurociencia cognitiva de primera clase necesita hoy un buen filósofo crítico y analítico. El diálogo sobre las metas y los logros de la neurociencia cognitiva reproducido en este libro, que tuvo lugar en el seno de la reunión de la APA de 2005 en Nueva York, confirma esa creencia.

La tarea de la neurociencia es entender el funcionamiento del sistema nervioso, contribuyendo así al diseño de estrategias para aliviar a la humanidad de la terrible carga de enfermedades como la demencia o la esquizofrenia. Los neurocientíficos, en el cumplimiento de su tarea, también esclarecen aquellos mecanismos del cerebro que deben funcionar normalmente para que podamos ejercer nuestras facultades psicológicas, tales como la percepción y la memoria.

Esta concepción de la neurociencia se opone a la que sostiene que la neurociencia tiene un único objetivo global, el de entender la *conciencia.*[1] Es interesante considerar en detalle esta sugerencia, como ejemplo de la necesidad de un análisis filosófico crítico de las neurociencias del tipo que he propuesto más arriba. Peter Hacker y yo dedicamos en nuestro libro más de cien páginas al tema de la *conciencia.* Al principio de nuestro análisis afirmamos que «un primer paso hacia la claridad es distinguir la conciencia *transitiva* de la *intransitiva.* La conciencia transitiva es la capacidad de ser consciente *de* una u otra cosa, o de ser consciente *de que* una u otra cosa son de un modo u otro. La conciencia intransitiva, por el contrario, no tiene objeto. Es la capacidad de estar consciente o despierto, frente a estar inconsciente o dormido» (*Philosophical Foundations of Neuroscience*, en adelante PFN, pág. 244). La pérdida de la conciencia intransitiva, por ejemplo cuando se está dormido, desvanecido o anestesiado, es objeto de una rica literatura neurocientífica. Por otro lado, existe mucha confusión en la literatura neurocientífica cuando se trata del estudio de las diversas formas de conciencia transitiva. Estas formas incluyen la conciencia perceptiva, la somática, la cinética y la afectiva, la conciencia de las propias motivaciones, la conciencia reflexiva, la conciencia de las propias acciones y la autoconciencia (PFN, págs. 248-252). Algunas de estas formas de conciencia transitiva son atencionales. Por ejemplo, la conciencia perceptiva de algo implica que la cosa de la que se es consciente atrae la atención de modo sostenido. Es un error suponer que la percepción de algo es, en tanto que percepción, una forma de conciencia transitiva o incluso que conlleva ser consciente de lo que uno percibe.

La investigación neurocientífica se ha consagrado de manera abrumadora sólo a una forma de lo que se interpreta (o malinterpreta) como conciencia perceptiva transitiva, en concreto la que in-

terviene en la percepción visual, en particular el fenómeno de la rivalidad binocular. Durante esta rivalidad, el observador capta dos imágenes incongruentes, cada una de las cuales afecta a un ojo distinto, pero percibe sólo una imagen a la vez. La imagen perceptivamente dominante y la suprimida se alternan cada pocos segundos. El trabajo experimental de Logothetis y sus colegas (Leopold y Logthetis 1999; Blake y Logothetis, 2002) permite entender esto. Mediante el uso de técnicas de condicionamiento operante, entrenaron a unos monos para que accionaran una palanca para indicar cuál de las dos imágenes monoculares en competencia es la dominante en un determinado momento (véase figura 8). Los picos de actividad de los potenciales de acción, registrados en células individuales de la corteza visual del mono, se pueden correlacionar con la respuesta perceptiva del animal de mover la palanca. Ello permite identificar las regiones corticales en las que la actividad neuronal se corresponde con la experiencia perceptiva. El recuadro de la izquierda de la figura 8 muestra el número de picos de actividad eléctrica registrados en una

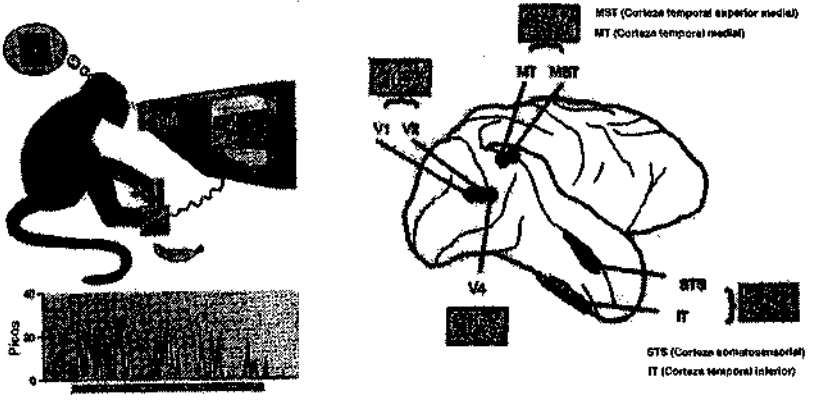

FIGURA 8: Técnicas de condicionamiento operante para establecer las áreas de la corteza con neuronas que, durante la presentación de imágenes monoculares «rivales», descargan picos de potenciales de acción correlacionados con los «informes perceptivos» de los monos (extraído de Blake y Logothetis, 2002; Leopold y Logothetis, 1999).

célula individual activa durante la rivalidad binocular. La barra bajo el eje x indica la percepción alterna de las dos imágenes, en clara correlación con los períodos de máxima actividad. La figura 8 (derecha) muestra las zonas del cerebro que contienen neuronas excitadas cuya actividad se correlaciona con la percepción visual del mono. El porcentaje de neuronas relacionadas con la percepción aumenta en los centros visuales «superiores», es decir, aquellas áreas de la corteza más alejadas del input procedente del tálamo. Sólo una pequeña fracción de las neuronas sensibles a estímulos visuales en las áreas corticales V1 (con conexiones directas con el tálamo) y V2, las primeras a la que llega el input talámico, respondían de acuerdo con las alternancias de la rivalidad binocular. En cambio, el porcentaje era superior en las áreas más alejadas del input talámico, concretamente: V4, corteza temporal medial (MT) y corteza temporal superior medial (MST). La actividad de casi todas las neuronas sensibles a estímulos visuales de las zonas de la corteza temporal inferior (IT) y somatosensorial (STS) coincidía en gran medida con el estado perceptivo del animal. Esta modulación de las neuronas corticales contrasta con la actividad de las neuronas no corticales del núcleo geniculado lateral, que tienen conexiones directas con la retina. Éstas no muestran ninguna modulación durante la rivalidad binocular.

Así pues, hay en la corteza un conjunto distribuido de neuronas que se activan en coordinación con las reacciones perceptivas del animal, si bien se activan menos neuronas en las áreas visuales primarias (V1) que en las superiores. Lumer y sus colegas, utilizando imágenes por RMF en sujetos humanos, han determinado la distribución espacial de las neuronas de la corteza que están activas durante esas reacciones a la alternancia binocular. Sus estudios revelan que las neuronas cuya actividad fluctúa en coordinación con las fluctuacio-

nes reportadas de la experiencia perceptiva están ampliamente distribuidas por la corteza, incluyendo la corteza prefrontal lateral y los centros visuales superiores (Lumer, Friston y Rees, 1998). Lo mismo observaron Edelman y sus colegas, que utilizaron magnetoencefalografía (MEG) para registrar la actividad cortical de sujetos humanos (Tononi y Edelman, 1998; Srinivasan y otros, 1999). Este tipo de actividad distribuida durante la percepción parece ser lo que ha llevado al profesor Searle a sugerir que existe un «campo de la conciencia».[2]

En contraste con el énfasis en la distribución de la actividad de la corteza durante la rivalidad binocular, otros autores señalan que sólo determinadas clases de neuronas de los centros visuales superiores pueden considerarse candidatas a correlato neuronal de la conciencia (CNC) para este fenómeno. Por ejemplo, Crick y Koch (2003) subrayan que sólo los centros visuales superiores, como STS e IT, poseen neuronas sensibles a estímulos visuales que reaccionan típicamente (90 %) en coordinación con las alternancias de la rivalidad binocular. Esto lleva a Crick y Koch a examinar los detalles de «las arborizaciones dendríticas de los diferentes tipos de neuronas del giro temporal inferior (IT) del macaco que se proyectan hacia la corteza prefrontal, cerca del surco principal» (fig. 9; arriba, destacada en gris). Y señalan que «sólo un tipo de células tienen dendritas apicales que alcanzan la capa 1». A continuación, preguntan: «¿Qué hay de especial en la actividad que supera el umbral de la conciencia? Podría ser la descarga de tipos especiales de neuronas, como las células piramidales que proyectan sus axones hacia la parte frontal del cerebro» (fig. 9). Es decir, ahora tenemos que los CNC de la percepción durante la rivalidad binocular se identifican con un determinado tipo de neurona de la corteza temporal inferior, una idea similar a la de una «célula pontífice» o las «células cardenal» de

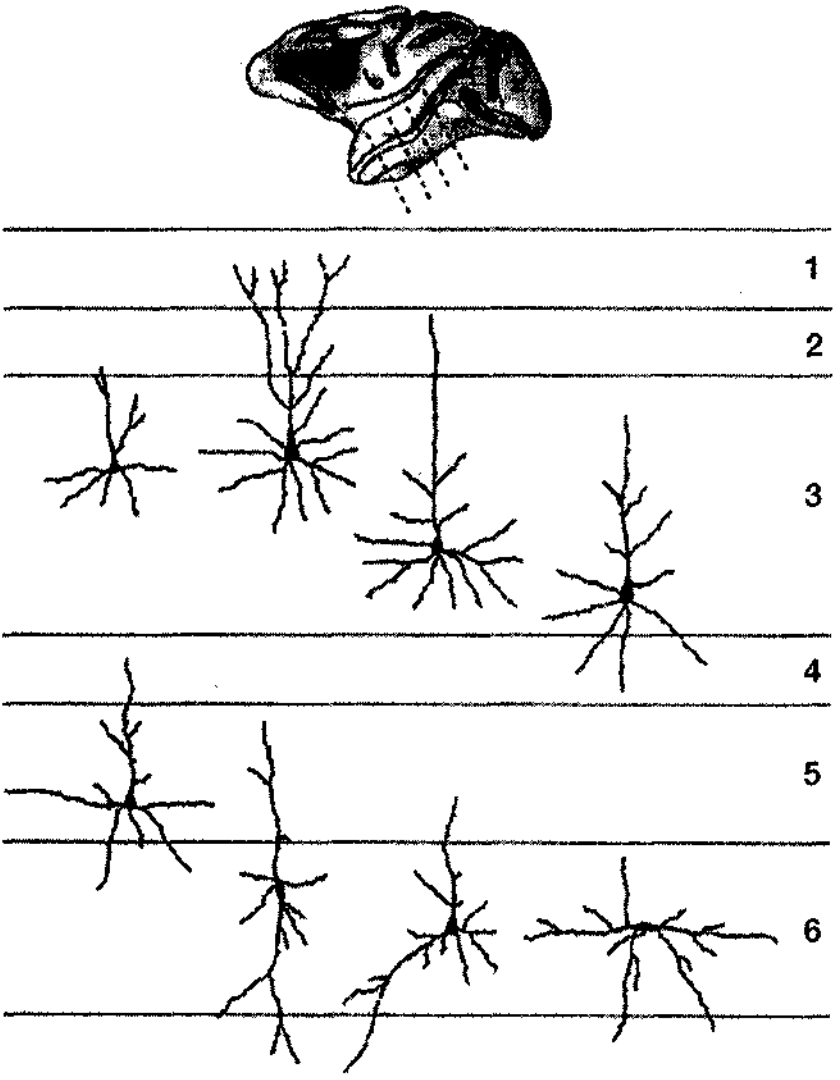

FIGURA 9: Collage de neuronas de la corteza temporal inferior (IT en la figura 8) que se proyectan hacia una porción limitada de la corteza prefrontal (zona punteada del cerebro representado en la parte superior; extraído de Crick y Koch, 2003, proveniente de De Lima, Voight y Morrison, 1990).

Barlow (1997).* Esto lleva a Koch (2004), en su libro *The Quest for Consciousness*, objeto de gran admiración por parte de venerables filósofos,[3] a describir mediante la figura 10 lo que la neurociencia habría supuestamente revelado respecto a los CNC durante la percepción visual. Se trata de una concepción criptocartesiana de la relación entre nuestros atributos psicológicos y el funcionamiento del cerebro. Una idea que hunde sus raíces en la falsa creencia, que señalamos en nuestro libro (PFN, cap. 10), de que la esencia de la conciencia es su asociación con los *qualia*, y de ahí que se pueda descubrir que estas inefables características cualitativas de la experiencia son causadas por células cardenales o pontífices.

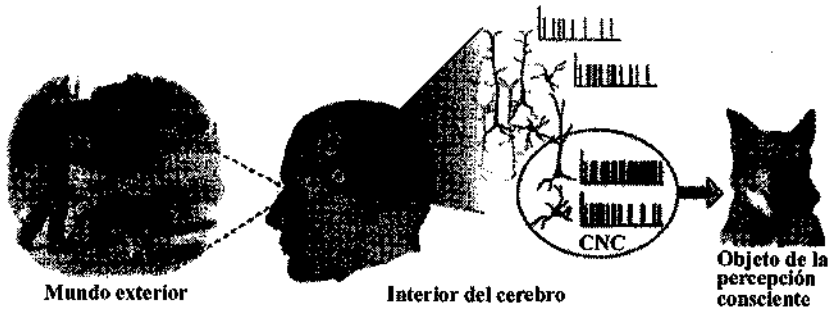

FIGURA 10: «Los CNC son el conjunto mínimo de acontecimientos neurales —aquí potenciales de acción sincronizados de las neuronas piramidales neocorticales (de la zona IT)— suficientes para una determinada percepción consciente» (extraído de Koch, 2004).

* Alusión a una metáfora de Sherrington (1941) en referencia a que los sistemas perceptivos corticales no podían estar organizados en jerarquías estáticas con una única neurona «pontífice» en la cúspide, retomada críticamente por Barlow para abogar por «asambleas cardenalicias», agregaciones dinámicas de células en lo alto de jerarquías igualmente estáticas, destacando la importancia para la percepción de la actividad de neuronas específicas. Véase Sherrington, C. S., *Man on his nature*, Cambridge, Cambridge University Press, 1941 (trad. cast.: *El hombre en su naturaleza*, Madrid, Alhambra, 1947. *(N. del t.)*

Creo que esta breve exposición de la investigación neurocientífica, y su interpretación de las percepciones visuales durante la rivalidad binocular, revela una necesidad apremiante de esclarecimiento crítico por parte de los filósofos analíticos. Peter Hacker y yo sugerimos que los resultados de esta investigación, tanto si se interpretan en términos de un «campo de la conciencia» como de «células cardenales», no contribuyen en nada a la comprensión de la conciencia perceptiva transitiva. En el mejor de los casos, contribuyen a la identificación de algunos de los correlatos neurales de la percepción visual en condiciones de rivalidad binocular. Pero la conciencia perceptiva transitiva implica que un objeto del campo visual atrae nuestra atención de forma sostenida. Para descubrir los CNC de la conciencia perceptiva transitiva, un estudio neurocientífico debe referirse a los correlatos neurales de tener la atención cautiva por lo que uno percibe, no de la percepción en *sí* misma. Porque percibir un objeto no es lo mismo que ser consciente del objeto que se percibe. Alguien puede percibir un objeto X sin ser consciente de ello, sea porque lo identifica erróneamente como Y, sea porque la atención no es atraída por él de forma sostenida, lo cual se podría deber a que ni siquiera lo percibe o a que se fija en él *intencionadamente*. La conciencia perceptiva transitiva es una forma de *receptividad* cognitiva (PFN, págs. 253-260). La conciencia afectiva, la conciencia de las propias motivaciones, la conciencia reflexiva de las propias acciones y la autoconciencia, exigen formas de análisis que difieren en aspectos importantes. En *Philosophical Foundations of Neuroscience* y en nuestro nuevo libro *History of Cognitive Neuroscience: A Conceptual Analysis*, de próxima aparición, se incluyen muchos otros ejemplos que demuestran la necesidad de esclarecimiento, y que abarcan en su totalidad el amplio abanico de investigaciones neurocientíficas sobre los correlatos neuronales de nuestros atributos psicológicos.

Si los filósofos se convierten en acólitos de la empresa neurocientífica, no harán más que amplificar el endiosamiento de la neurociencia que mencionaba en la introducción. Lo que la disciplina necesita no son reseñas laudatorias en el *New York Review of Books* sobre las obras de neurocientíficos aparentemente ignorantes de las dificultades conceptuales manifiestas en sus ideas. Lo que nos hace falta es una crítica filosófica iluminadora que pueda guiar una investigación neurocientífica fructífera sobre nuestras capacidades psicológicas y su ejercicio. Es ésta, en mi opinión, una tarea importante para las nuevas generaciones de filósofos.

Referencias bibliográficas

Barlow, H. B. y W. R. Levick, «The mechanism of directionally selective units in rabbit's retina», en *Journal of Physiology*, nº 178, 1965, págs. 477-504.

Bennett, M. R., *History of the Synapse*, Amsterdam, Harwood Academic, 2001.

Bennett, M. R., L. Farnell y W. G. Gibson, «A quantitative model of purinergic junctional transmission of calcium waves in astrocyte networks», en *Biophysics Journal*, nº 89, 2005, págs. 2.235-2.250.

Bennett, M. R., W. G. Gibson y J. Robinson, «Dynamics of the CA3 pyramidal neuron autoassociative memory network in the hippocampus», en *Philosophical Transactions of the Royal Society of London B Biological Sciences*, nº 343, 1994, págs. 167-187.

Bennett, M. R. y P. M. S. Hacker, *Philosophical Foundations of Neuroscience*, Oxford, Blackwell, 2003.

Bennett, M. R. y A. G. Pettigrew, «The formation of neuromuscular synapses», en *Cold Spring Harbor Symposium in Quantitative Biology*, nº 40, 1975, págs. 409-424.

Blakemore, C., *Mechanisms of the Mind*, Cambridge, Cambridge University Press, 1997.

Bridley, G. S., «The classification of modifiable synapses and their use in models for conditioning», en *Proceedings of the Royal Society London B*, nº 168, 1967, págs. 361-376.

—, «Nerve net models of plausible size that perform many simple learning tasks», en *Proceedings of the Royal Society London B*, nº 174, 1969, págs. 173-191.

Brodman, K., *Vergleichende Lokalisation: lehre der Grosshirnrinde in ihren' Prinzipien dargestellt auf Grund des Zellenbaues*, Leipzig, Barth, 1909.

Cornell-Bell, A. H., S. M. Finkbeiner, M. S. Cooper y S. J. Smith, «Glutamate induces calcium waves in cultured astrocytes: long-range glial signalling», en *Science*, nº 247, 1990, págs. 470-473.

Fried, S. I., T. A. Munch y F. S. Werblin, «Mechanisms and circuitry underlying directional selectivity in the retina», en *Nature*, nº 420, 2002, págs. 411-414.

—, «Directional selectivity is formed at multiple levels by laterally offset inhibition in the rabbit retina», en *Neuron*, nº 46, 2005, págs. 117-127.

Gazzaniga, M. S., R. B. Ivry y G. R. Mangun, *Cognitive Neuroscience: The Biology of the Mind*, 2ª ed., Nueva York, Norton, 2002.

Giannikopoulos, D. V. y U. T. Eysel, «Dynamics and specificity of cortical map reorganization alter retinal lesions», en *Proceedings of the National Academy of Sciences, USA*, nº 103, 2006, págs. 1.085-1.010.

Gibson, W. G., J. Robinson y M. R. Bennett, «Probabilistic secretion of quanta in the central nervous system: Granule cell synaptic control of pattern separation and activity regulation», en *Philosophical Transactions of the Royal Society of London B Biological Sciences*, nº 332, 1991, págs. 199-220.

Gilbert, C. D., «Adult cortical dynamics», en *Physiology Review*, nº 78, 1998, págs. 467-485.

Gilbert, C. D., J. A. Hirsch y T. N. Wiesel, «Lateral interactions in vi-

sual cortex», en *Cold Spring Harbor Symposium in Quantitative Biology*, nº 55, 1990, págs. 663-677.

Gilbert, C. D. y T. N. Wiesel, «Receptive field dynamics in adult primary visual cortex», en *Nature*, nº 356, 1992, págs. 150-152.

Hubel, D. H., T. N. Wiesel y S. LeVay, «Functional architecture of area 17 in normal and monocularly deprived macaque monkeys», en *Cold Spring Harbor Symposium in Quantitative Biology*, nº 40, 1976, págs. 581-589.

Ito, M., «Cerebellar long-term depression: Characterization, signal transduction, and functional roles», en *Physiology Review*, nº 81, 2001, págs. 1.143-1.195.

—, «The molecular organization of cerebellar long-term depression», en *Nature Reviews Neuroscience*, nº 3, 2002, págs. 896-902.

Ito, M. y M. Kano, «Long-lasting depression of parallel fiber: Purkinje cell transmission induced by conjunctive stimulation of para llel fibers and climbing fibers in the cerebellar cortex», en *Neuroscience Letters*, nº 33, 1982, págs. 253-258.

Koch, C., *The Quest for Consciousness*, CO, Roberts, 2004.

Lalonde, R., A. N. Bensoula y N. Filali, «Rotorod sensorimotor learning in cerebellar mutant mice», en *Neuroscience Research*, nº 22, 1995, págs. 423-426.

McCormick, D. A. y R. F. Thompson, «Cerebellum: Essential involvement in the classically conditioned eyelid response», en *Science*, nº 223, 1984, págs. 296-299.

Marr, D., «A theory of cerebellar cortex», en *Journal of Physiology*, nº 202, 1969, págs. 437-470.

—, «Simple memory: A theory for archicortex», en *Philosophical Transactions of the Royal Society of London B Biological Sciences*, nº 262, 1971, págs. 23-81.

Masland, R. H., «The fundamental plan of the retina», en *Nature Neuroscience*, nº 4, 2001, págs. 877-886.

Polyak, S. L., *The Retina*, Chicago, University of Chicago Press, 1941.

Ramon y Cajal, S., *Histology of the Nervous System* (1904), Oxford, Oxford University Press, 1995 (ed. cast.: *Histología del sistema nervioso del hombre y de los vertebrados*, Madrid, Consejo Superior de Investigaciones Científicas, 1911).

Sieburth, D., O. Ch'ng, M. Dybbs, M. Tavazoie, S. Kennedy, D. Wang, D. Dupuy, J. F. Rual, D. E. Hill, M. Vidal, G. Ruvkun y J. M. Kapjan, «Systematic analysis of genes required for synapse structure and function», en *Nature*, nº 436, 2005, págs. 510-517.

Smirnakis, S. M., A. A. Brewer, M. C. Schmid, A. S. Tolias, A. Schuz, M. Augath, W. Inhoffen, B. A. Wandell y N. K. Logothetis, «Lack of long-term cortical reorganization after macaque retinal lesions», en *Nature*, nº 435, 2005, págs. 300-307.

Vaney, D. I. y W. R. Taylor, «Direction selectivity in the retina», en *Current Opinion in Neurobiology*, nº 12, 2002, págs. 405-410.

Welsh, J. P., H. Yamaguchi, X. H. Zeng, M. Kojo, Y. Nakada, A. Takagi, M. Sugimori y R. R. Llinas, «Normal learning during pharmacological prevention of Purkinje cell long-term depression», en *Proceedings of the National Academy of Sciences, USA*, nº 102, 2005, págs. 17.166-17.171.

Young, J. Z., *Programs of the Brain*, Oxford, Oxford University Press, 1978.

Zeki, S., «Splendours and miseries of the brain», en *Philosophical Transactions of the Royal Society of London B Biological Sciences*, nº 354, 1999, págs. 2.053-2.065.

LA BÚSQUEDA CONTINÚA

La ciencia y la filosofía en busca del Príncipe Razón

DANIEL ROBINSON

En el ambiente intelectual de la Gran Bretaña de principios del siglo XVII se respiraba un gran interés por la anatomía. El gran William Harvey, que había regresado de Italia en 1602, impartiría sus pioneras lecciones Lumley a partir de 1615. Cambridge fue uno de los centros de este interés renovado por la maquinaria del cuerpo. Harvey había salido de Cambridge en 1600, título en mano, para recibir instrucción en Padua por parte del propio Fabricius, mientras su joven colega Phineas Fletcher (1582-1650) estaba completando su propia carrera académica en el King's College. Todos sabemos quién era a Harvey, pero Fletcher yace casi olvidado en las brumas del pasado. Ningún estudiante de anatomía cambiaría el modo de instrucción de Fletcher por el de Harvey. Sin embargo, la curiosidad que nos lleva a revivir el enfoque de Fletcher se ve recompensada con creces, en tanto que la forma en que comprendemos el mundo natural, o parte de él, está íntimamente ligada a los métodos seleccionados para la tarea.

Publicada en 1633, *The Purple Island,* de Phineas Fletcher, es una alegoría en doce cantos que guían al lector por el misterioso territorio del cuerpo humano. La isla debe su color a la materia de color púrpura con la que Dios creó la nueva tierra. La canción del

descubrimiento anatómico es cantada por Thirsil a un público de jóvenes pastores. Si imaginamos que esos mozos se parecen a los estudiantes de hoy, convendremos en que su interés no se despierta hasta que Thirsil llega al canto 6, pues en la estrofa 28 entramos en el reino de

> El Príncipe de las Islas, de constitución más que celestial,
> [...] con razón llamado el *Intelecto* que todo lo ve;
> Todo de un brillante glorioso, pues nada es terrestre;
> Cuya cara se asemeja al sol, y el más divino aspecto
> Ninguna visión humana podrá describir jamás:
> Pues cuando a él uno dirige los ojos,
> Se sume en la torpeza o la sorpresa ante tan gran majestad.

Después, al llegar a la estrofa 30, no queda duda alguna sobre la composición del Príncipe:

> Su cuerpo tan extraño no es corpóreo,
> Sino materia sin materia; nunca lleno,
> Ni llenándose; aunque tiene a su alcance
> Todo el cielo y toda la tierra y todo lo que en ellos hay;
> Pero podría contener miles de miles de cielos,
> Y seguir tan vacío como al principio;
> Y cuanto más toma, más dispuesto está para tomar más.[1]

Harvey y Fletcher habían emprendido la antigua búsqueda de esa sede del alma racional, de ese «lugar de las formas» que Aristóteles sabiamente no se propuso localizar, de ese «cuerpo extrañísimo» cuyas propiedades definitorias parecen ser cualquier cosa menos corporales. Como deja claro el presente volumen, la búsqueda continúa, y como deja aún más claro, hay menos certeza hoy que en el

pasado sobre qué es lo que probablemente se descubrirá y en qué lugares es más probable que se descubra.

Entre los coautores de este volumen hay un destacado científico y un grupo de filósofos consumados e influyentes. Con el fin de situar las diferencias que animan sus intercambios en el contexto más amplio de la historia de las ideas, es útil considerar una vez más los métodos de Harvey y Fletcher, ambos emprendiendo un viaje de descubrimiento, ambos comprometidos con un modo de explicación aparentemente revalidado por prácticas humanas de valor demostrado. Visto bajo esa luz, y con las reservas que debidamente señalaré más adelante, diría que John Searle y Daniel Dennett desearían ser asociados a Harvey, pese a que sus filosofías especulativas forman parte en realidad de la tradición de Fletcher. Al igual que éste, son estudiantes de la anatomía de su época que la emplean para contar una historia. Del lector depende interpretar sus conclusiones como alegorías o como titulares periodísticos. Sin embargo, en tanto que relatos, no deben confundirse con la bien distinta misión de la ciencia, tanto la experimental como la teórica.

Max Bennett y Peter Hacker, al llamar hábilmente la atención sobre esto, insistirían con tenacidad en que forma parte de la propia naturaleza del caso, en que son caminos distintos que conducen hacia metas respetables pero fundamentalmente distintas. Dejando aquí de lado cualquier juicio sobre el valor relativo de los distintos modos de indagación y de explicación, podríamos señalar que las páginas de Bennett y Hacker (a excepción de aquellos pasajes explícitamente técnicos de Bennett) hunden sus raíces en las tradiciones largo tiempo aceptadas de la filosofía analítica. Con esto no me refiero a cierto presunto «descubrimiento» por parte de un atajo de filósofos oxonienses de razonamiento simplista, sino a esa misión central de los diálogos de Platón —el esclarecimientos de los términos, el plan-

teamiento de los problemas en forma argumentativa, la exigencia de coherencia y consistencia—. A todo esto Aristóteles añadiría el contenido del mundo natural y el ensanchamiento de la misión filosófica que se deriva de ese añadido. Dennett y Searle tienen historias que contar, y son buenas historias contadas por maestros en el oficio. Bennett y Hacker, en su importante libro *Philosophical Foundations of Neuroscience*, concluían que el valor de verdad de tales historias no se podía evaluar, debido a que la elección de los términos es peculiar, acientífica y ajena a la filosofía. Lo que encontraron en Searle, Dennett y otros muchos líderes actuales del pensamiento de la ciencia cognitiva, fue simplemente un exceso de lo que Gilbert Ryle aludía irónicamente con la frase «ella llegó en un mar de lágrimas y en palanquín».* En este libro, se remite al lector a esas historias en forma condensada e instructivamente dialéctica.

Al reflexionar en su propio capítulo, Bennett relata un encuentro con John Eccles, quien socarronamente insistía en que la investigación en neurología siempre requiere «un buen soldador». Bennett expresa entonces su propia convicción de que «todo laboratorio de neurociencia cognitiva de primera clase necesita hoy un buen filósofo crítico y analítico». Después de dedicar muchos años tanto al laboratorio como a la reflexión, sé de primera mano que Eccles estaba en lo cierto, pero no estoy tan seguro de que Bennett lo esté. Mi duda surge de un escepticismo más generalizado sobre la idoneidad de combinar con o sin guión disciplinas bien definidas. Una vez que una disciplina bien definida como la ética se recombina en algo llamado bioética, hay una tendencia a pensar que se debe buscar al-

* Alusión al gusto por la ornamentación retórica, el dramatismo y la superficialidad por parte de la ortodoxia intelectualista en la filosofía de la mente contemporánea. *(N. del t.)*

gún precepto ético más profundo que cubra el caso de que a uno le roben el hígado, ya que los preceptos que cubren el robo de un coche resultan insuficientes para esa tarea. Por su parte, en el compuesto «neurociencia cognitiva» el propio adjetivo plantea un tema refractario al análisis filosófico, al menos desde los días en que Platón dio voz al Sócrates del *Critón*. Lo primero que se pregunta Bennett —cómo derivar lo psicológico de lo sináptico— es también la pregunta de Fletcher. Lo que a éste le faltaba, en su intento por *ver* al Príncipe, no se lo podía proporcionar Harvey. Tampoco sirven aquí ni el bisturí ni el laparoscopio. Existen sólidos argumentos para pensar que la misión ensanchada de una *psicología cognitiva* realista e informada debe discurrir por un camino distinto del que guía el progreso del científico como peregrino.

Tal vez pueda perfilar mejor mi idea aludiendo a la propia investigación que Bennett resume, para concluir que los resultados de la percepción visual están más correlacionados con procesos corticales «superiores» que con los más próximos a su origen en la retina. Hay una serie larga y consistente de descubrimientos que establecen que, al menos en la audición y la visión, la «sintonización» del sistema se hace cada vez más precisa a medida que los acontecimientos pasan del nivel de las neuronas primarias a sus destinos corticales definitivos. Pero el fenómeno de la rivalidad binocular es diferente de la sintonización del sistema en bandas de frecuencia más estrechas. Hay una evidente propiedad fenomenológica que la conducta condicionada del animal señala la primera vez que una imagen y luego la otra se alternan en la dominancia perceptiva. La pregunta que naturalmente surge dentro de este marco tiene que ver con la forma en que todo esto tiene lugar en criaturas con los ojos dispuestos lateralmente, de modo que no puede existir el mismo tipo de *rivalidad* binocular. Los animales con ojos situados medialmente se enfrentan a un espacio

visual en el que el mismo objeto puede competir por el reconocimiento. Los que tienen los ojos situados a los lados se enfrentan a dos campos visuales separados que no tienen ningún objeto en común.

¿Por qué menciono esto? Lo hago para señalar la idea obvia (aunque normalmente pasada por alto) de que no sólo hay un *animal* real que está viendo algo, sino que la forma en que el entorno visible influye en su percepción depende de una realidad ecológica mucho más amplia y compleja, a la que las criaturas de una determinada clase deben adaptarse. Este hecho impone límites de diversos grados de severidad a las generalizaciones entre las especies, y no parece descabellado sospechar que, cuando esas generalizaciones incluyen la conciencia «transitiva» o la «intransitiva», entran en juego límites aún más estrictos. Del mismo modo que el pez no puede descubrir el agua, las observaciones realizadas en la burbuja artificial que es un laboratorio resultan irremediablemente alejadas de la vida real vivida en el mundo visible. Dicho de otra forma, resulta muchísimo más fácil determinar cómo se siente al ser un gato que determinar cómo se siente al no haber visto nada más que proyecciones diseñadas para excitar las células retinales, durante toda una vida pasada en una jaula del laboratorio.

Entre los recelos expresados por Bennett y Hacker ante neurocientíficos cognitivos por otro lado libres de sospecha, destaca especialmente el que tiene que ver con la afirmación de que no siempre somos conscientes de lo que percibimos. Bennett observa que «alguien puede percibir un objeto X sin ser consciente de ello, sea porque lo identifica erróneamente como Y, sea porque la atención no es atraída por él de forma sostenida, lo cual se podría deber a que ni siquiera lo percibe o a que se fija en él *intencionadamente*». Esto no resulta muy convincente. Que se identifique erróneamente un objeto no implica en absoluto que no se haya percibido conscientemente

algo. La «identidad» de cualquier objeto visual no se establece unívocamente. El amor que Edipo sentía por Yocasta no era filial. La flor azul, que a la luz de la luna brilla más que la amarilla, que a su vez brilla más al mediodía, no es una prueba en contra de la conciencia perceptiva transitiva; tampoco lo sería el testimonio de la abeja, cuya máxima sensibilidad espectral está en la banda ultravioleta del espectro electromagnético. Además, un índice fiable de la actuación de la conciencia perceptiva transitiva es la segregación *intencionada* de ítems en el espacio visual. Por lo que se refiere a percibir X «sin ser consciente de ello», me temo que un consenso generalizado sobre esta cuestión requeriría algún tipo de defensa teórica especial.

Son éstos escrúpulos más bien menores, en especial si se los compara con los que plantea John Searle. El objetivo de éste es situar de nuevo la conciencia en el cerebro, que es donde la ubicaban hace mucho tiempo Galeno y los hipocráticos. Searle supera a éstos en sus conclusiones, beneficiándose del progreso científico y por medio de la distinción entre *estado* y *lugar*. Se presume que la afirmación de que la conciencia es un *estado* del cerebro es menos peculiar que la de que está «en» el cerebro.

Se ha recurrido con tanta frecuencia a hablar de estados y términos emparentados como procesos y mecanismos, que tales términos gozan hoy de un estatus casi protegido. Pero no está del todo claro que tengan otro propósito que el de imponer de forma ilícita conclusiones a una argumentación que propiamente aún no se ha hecho. Y se trata de una práctica que crea hábito. Searle se permite postular «estados» e inmediatamente les añade algo llamado «el carácter cualitativo de beber cerveza» que, según dice, «es diferente del carácter cualitativo de escuchar la *Novena Sinfonía* de Beethoven». Estoy razonablemente seguro de que conozco la diferencia entre beber cerveza y escuchar a Beethoven, pero estoy mucho menos segu-

ro de que ninguna de esas actividades tenga un «carácter cualitativo». Nunca he consumido un carácter cualitativo, aunque he tomado bastante cerveza. ¿Es ésta una objeción nimia? Tal vez. Pero, en este campo, las objeciones nimias son acumulativas y pueden alcanzar el equivalente filosófico de una masa crítica.

Searle llega al núcleo de su crítica con su admirable concisión característica:

> Gran parte de los mejores trabajos en neurociencia constituyen un esfuerzo por explicar de qué forma los procesos cerebrales causan la experiencia visual y dónde y cómo se produce ésta en el cerebro. Asombrosamente, Bennett y Hacker niegan la existencia de la experiencia visual en este sentido, en el sentido de *quale*.

Antes de considerar la asombrosa negación de Bennett y Hacker, es importante examinar la que yo estimo es una afirmación mucho más asombrosa, la de que los mejores trabajos en neurociencia prometen explicar de qué forma los procesos visuales *causan* la experiencia visual (que se produce en alguna parte del cerebro). John Searle sabe perfectamente que todo este tema de la causalidad es esencial para el asunto que nos ocupa. Sabe, por consiguiente, que la afirmación de que tal causalidad está bien establecida —basándose en sólidos estudios que demuestran cómo funciona todo— no resiste el análisis filosófico. No existe acuerdo unánime sobre cuáles son los propios términos de la relación causal (¿se trata de hechos, objetos materiales, términos conceptuales, acontecimientos, condiciones?), ni siquiera sobre si deben existir invariablemente. Al fin y al cabo, la causa de que Juana sobreviviera fue el hecho de que no se bebió el veneno. Aquí la «causa» de la supervivencia es un no-acontecimiento. Para ir al grano, si todo lo que define el dominio de lo

mental es en efecto el resultado causal de algunos conjuntos de «estados» del cerebro, entonces, como reza la máxima, *la física como ciencia está completa* y los filósofos deberían considerar reciclarse en otra carrera profesional.

Como residente desde hace muchos años del planeta Tierra, tengo pocas dudas de que una organización sana y funcional del cuerpo, en especial si se incluye el sistema nervioso, constituye la *condición* necesaria para lo que nos complacemos en llamar nuestra vida mental, al menos en sus encarnaciones sublunares. No obstante, la sugerencia de que el tejido nervioso *causa* todo esto despertaría cuando menos estupefacción en una época en que la ciencia no se hubiera convertido aún en una especie de retórica. Justo en el momento en que los pensadores punteros en el campo de la física se muestran disciplinadamente cautelosos en lo que a la causalidad se refiere, ahí llegan los neurocientíficos cognitivos y sus lacayos filosóficos preguntándose cómo alguien puede dudar en adoptar un objetivo tan obvio: «de qué forma los procesos cerebrales causan la experiencia visual y dónde y cómo se produce ésta en el cerebro». Aun suponiendo que pudiéramos afirmar, en algún sentido metafísicamente aceptable, que hemos establecido de qué forma las fuerzas gravitatorias *causan* que la llave de la puerta caiga hacia el centro de la Tierra, la explicación funciona (si es que lo hace en absoluto) sólo en la medida en que la Tierra y las llaves de casa tienen ambas masa, y la magnitud de su separación se puede especificar en millas, pies, pulgadas o (a regañadientes) metros. Sin embargo, el listón metafísico se coloca mucho más alto cuando la conexión causal es la que hay entre la actividad metabólica *en algún lugar* por un lado y escuchar el *lied* «An die Freude», y no digamos componerlo, por el otro.

Searle manifiesta serias dudas sobre el alcance del juego del lenguaje. Los recursos conceptuales que Bennett y Hacker emplean pro-

fusamente son tomados de Wittgenstein, y Searle plantea la que de hecho es una reserva bien conocida ante, en sus palabras, el error wittgensteiniano de confundir

> los criterios de la aplicación de conceptos mentales con los propios estados mentales. Es decir, confunden los criterios conductuales de la *adscripción* de predicados psicológicos con los *hechos adscritos por estos* predicados psicológicos, lo cual es un grave error.

La cuestión es demasiado abstrusa como para abordarla sumariamente. No se puede discutir que las explicaciones del dolor hechas en primera y tercera persona se obtienen de fuentes distintas. Es indiscutible que la base sobre la que Pérez *siente* dolor de muelas es diferente de la base sobre la que Martínez juzga que Pérez sufre dolor. Sin embargo, lo que provoca polémica es la afirmación de que la mera presencia de patrones de descarga excesivos en las fibras relevantes de la rama maxilar del nervio trigémino justifican que se pueda decir que Pérez siente «dolor» en exactamente el mismo sentido en que lo dirían personas social y lingüísticamente competentes en su cultura. Tengo mis dudas al respecto pero, una vez más, la falta de espacio me impide una exposición más completa. Sin embargo, Searle, después de señalar la deuda con Wittgenstein, toma un pasaje de Bennett y Hacker sólo para interpretarlo de forma peculiar. Éste es el pasaje:

> Los criterios de la atribución de un predicado psicológico son [...] en parte constitutivos del *significado* de ese predicado [...] El cerebro no cumple los criterios para ser un posible sujeto de predicados psicológicos.
>
> (pág. 83)

Searle interpreta estas palabras como una negación de que el cerebro tenga conciencia en base a que el cerebro no puede «comportarse» de ningún modo. Dice que la tesis principal de Bennett y Hacker es que «el cerebro es incapaz de exhibir la adecuada conducta». Efectivamente, esto es lo que Bennett y Hacker dicen. Sin embargo, Searle pasa por alto el sutil (quizá demasiado sutil) argumento del que extraen esa conclusión. No es que no pueda atribuirse conciencia al cerebro porque éste es incapaz de exhibir la adecuada conducta, lo que ocurre es que las adscripciones en cuestión, para ser significativas en el sentido estricto de la palabra, deben cumplir los mismos criterios que debe cumplir cualquier predicado. Las afirmaciones de que Pérez es alto, el cerebro es húmedo y Enriqueta es joven son inteligibles sólo en la medida en que «húmedo», «joven» y «alto» no se saquen de la caja con la etiqueta ESCARABAJO, que además es visible sólo para quien la sostiene. El sentido en que Pérez, como individuo aislado, podría no atribuir significado al adjetivo «alto» con que los demás le califican es el mismo sentido en que podría decir impropiamente que algo «duele» *incluso hablando de sí mismo*. A no ser que también yo haga una lectura errónea de Bennett, Hacker y Wittgenstein, la conclusión no es que el cerebro no puede ser consciente, sino que la afirmación de que lo es resulta tan incomprensible como la de que es socialdemócrata.

En diversos momentos, pero sobre todo hacia el final de su interesante ensayo, Searle alude a las posibles aportaciones de la ciencia experimental a los problemas filosóficos. Señala que no es previsible que las cuestiones de «la vida buena», etc, reciban tal beneficio, no obstante espera que algunos problemas filosóficos sí sucumban ante los descubrimientos científicos. Es una lástima que el ejemplo concreto que ofrece pueda dejar irremediablemente perplejo a más de un lector. Éstas son las palabras de Searle:

> no hago una distinción nítida entre cuestiones científicas y cuestiones filosóficas. Permítaseme poner un ejemplo para explicar de qué forma los descubrimientos científicos pueden ayudar a mi trabajo filosófico. Cuando levanto el brazo, mi intención-en-acción consciente causa un movimiento físico de mi cuerpo. Pero el movimiento también tiene otro nivel de descripción, porque es causado por una secuencia de descargas neuronales y la secreción de acetilcolina en las placas terminales de los axones de las motoneuronas. Partiendo de estos hechos, puedo hacer un análisis filosófico para demostrar que uno y el mismo acontecimiento debe ser a la vez un acontecimiento consciente cualitativo y subjetivo, y poseer también muchas propiedades químicas y eléctricas. Pero ahí termina el análisis filosófico. Ahora necesito conocer el mecanismo exacto que hace que esto funcione.

Al leer esto, uno puede preguntarse qué tipo de «análisis filosófico» llega a la conclusión de que el movimiento del brazo tiene «muchas propiedades químicas y eléctricas». En un nivel, es evidente (y lo habría sido para los iliteratos habitantes de las cavernas) que los brazos pesan, que algo bajo la piel del brazo se tensa y que esos mismos brazos, cuando en ellos penetran objetos punzantes, exudan un líquido rojo y caliente. No hay discusión en que corresponde a la ciencia experimental averiguar los detalles de todos los acontecimientos subcutáneos asociados con levantar un brazo. Es dudoso que un análisis filosófico fuera algo más que una distracción mientras el equipo de investigación se emplea a fondo en este importante trabajo. Entretanto, quienes tengan inclinaciones filosóficas se podrían preguntar simplemente cuál es la diferencia entre el hecho de que a alguien le levanten el brazo y el de que alcance el mismo resultado levantando el brazo *intencionadamente*. Resulta plausible concluir, sin necesidad de ningún tipo de investigación científica, que una diferencia de este tipo se expresará en algún punto bajo la

piel, aún reconociendo al mismo tiempo que las diferencias psicoquímicas no «explicarán» las intenciones. Pero cuando aceptamos el cauteloso, fundamentado e informativo juicio de que los filósofos se han consagrado a la abstrusa cuestión de la actividad volitiva —todo ello mucho antes de que alguien supiera que existían las neuronas o los potenciales de placa terminal—, la proposición de que esos autores necesitaban conocer «el mecanismo exacto que hace que esto funcione» resulta poco creíble.

Searle está dispuesto a admitir las conclusiones de Wittgenstein en la cuestión del «juego del lenguaje», pero las considera irrelevantes para el proyecto de la neurociencia cognitiva. Así,

> cuando investigamos la ontología del dolor —no las condiciones para jugar al juego del lenguaje, sino la propia ontología del fenómeno en sí— podemos olvidarnos de la conducta exterior y limitarnos a averiguar de qué forma el cerebro causa las sensaciones internas.

Sin olvidar las mencionadas dudas acerca de «cómo el cerebro causa [...] las sensaciones», consideremos el sentido en que Searle habla de «ontología del dolor», refiriéndose a la sensación «real» misma. Un cerebro vivo nunca está inactivo, y por consiguiente el número de correlatos neurales-fenoménicos es efectivamente ilimitado. Las denominadas vías del dolor clásicas terminan en el tálamo, sin que exista un «centro» cortical del dolor como tal. Por lo tanto, la cuestión se reduciría entonces a cómo los núcleos talámicos «causan» el dolor. Los núcleos son constelaciones de cuerpos celulares integrados que funcionan como una unidad. De modo que podemos refinar aún más la búsqueda: ¿cómo causan el dolor los potenciales de acción postsinápticos emitidos por los cuerpos celulares del interior del tálamo?

Pongamos por ejemplo que estiramos o torcemos el brazo de alguien hasta el punto de que sienta dolor. El dolor, por supuesto, se siente en el brazo, no en el tálamo, porque no hay nada en el cerebro que «sienta» nada. Sabemos que Pepe sufre dolor, porque hace una mueca y dice: «¡Ay!». Sabemos también que las fibras C se activan, y que las señales que entran en la superficie dorsal de la médula espinal viajarán hasta el cerebro y los núcleos talámicos relevantes. Pero ese mismo viaje lo hacen muchas más cosas, procedentes del mismo brazo. Además, el «¡ay!» no se produce hasta que los niveles de descarga alcanzan y superan un valor crítico. Sin una reacción facial, postural o vocal por parte de Pepe, tendríamos todos estos datos neurofisiológicos, pero no sabríamos qué hacer con ellos. En última instancia, es Pepe quien tiene la última palabra sobre su dolor. ¿Pero cómo sabemos cuál de las «señales» de Pepe es la del dolor? ¿Y cómo lo sabe Pepe? Creo que Searle ha abandonado el juego del lenguaje cuando menos un fonema antes de tiempo.

En su discusión de la falacia mereológica, Searle rechaza que exista falacia alguna y sostiene que ni siquiera los argumentos wittgensteinianos pueden ahorrarse la noción de que (en un sentido especial) «el cerebro piensa» o «el cerebro ve». En palabras del propio Searle, Bennett y Hacker sostienen que el cerebro no piensa y que el pensamiento no puede tener lugar en el cerebro. Insiste en que «para demostrar que el cerebro no puede ser el lugar de tales procesos precisarían un argumento independiente, un argumento que no he podido encontrar». Ocurre más bien que

> Todo lo que el argumento wittgensteiniano exige es que el cerebro sea parte de un mecanismo causal de un sistema global capaz de producir la conducta. Y esta condición se puede cumplir incluso cuando determinados procesos psicológicos se localizan en el cerebro.

A medida que las objeciones nimias se van acumulando, uno empieza a percibir que hablar del pensamiento como un «proceso» exige más o menos que se le busque «su» lugar, y que sólo un argumento independiente en contra debilitará lo que dicta el sentido común en este punto. El pensamiento como «proceso» presumiblemente sería, por supuesto, un «proceso cerebral», al menos si hay que escoger entre los distintos órganos del cuerpo. Sin embargo, si «pensamiento» es la palabra que aplicamos a esa enorme cantidad de ocurrencias, expectativas, creencias, juicios, estrategias, etc., comprimidas en un implacable minuto de conciencia —o si la aplicamos a una de esas ideas con las que nos obsesionamos ininterrumpidamente durante largos periodos de tiempo—, diría que el peso de la prueba recae en aquellos que defienden que cualquiera de estas cosas es un «proceso», y más aún que tiene lugar en el cerebro. Pensemos en la *información*, esa entidad carente de masa y que no ocupa espacio, en el sentido estricto de aquello que altera las probabilidades y la entropía general de un sistema. No es éste el momento de entrar en el apasionante mundo de la indeterminación y la superposición cuánticas e intentar rescatar al gato de Schroedinger, pero resulta sin embargo provechoso reconocer que nuestra ciencia más desarrollada está mucho menos comprometida con la idea de que los efectos reales requieren lugares, masas y «procesos» observables. Se diría que el estatus definitivo del fisicismo depende precisamente de cuán mejor y más exhaustiva sea la explicación de la vida mental, pero seguramente es demasiado pronto para adoptar una posición firme sobre el particular. El punto de partida correcto es empezar por nuestra propia elección de términos, asegurándonos de que no adoptamos unos modos de hablar que virtualmente excluyen la posibilidad de desterrar nuestra ignorancia sistemática.

Daniel Dennett, con quien John Searle ha tenido sus propias discrepancias, es, no obstante, tan crítico con Bennett y Hacker como Searle, pero por razones distintas cuando no totalmente dispares. La principal línea de defensa de Dennett es presentarse a sí mismo como alguien que realmente extiende el pensamiento de «san Ludwig» (como le llama Dennett), aplicándolo a la conducta de robots, ordenadores que juegan al ajedrez e incluso cerebros y sus partes —conductas lo bastante similares a la de las personas para que sea posible la predicación de términos psicológicos—. Citando sus propios primeros escritos Dennett sostiene que, precisamente porque hay dos «niveles de explicación», estamos llamados a la tarea de relacionarlos, una tarea que requiere del análisis filosófico (*Content and Consciousness*, págs. 95-96). Sin embargo, el hecho de que haya en efecto dos niveles de explicación no es en sí mismo garantía de que estén o puedan estar relacionados ni, si lo están, de que esa relación sea de tipo causal. Si la relación que hubiese que examinar es la que se da entre, por ejemplo, la temperatura ambiente y la energía cinética media del sistema, tendríamos una relación de identidad. Pero si la relación es la que hay entre el número de una calle y la residencia de una determinada familia, es evidente que el dato que los Pérez viven en el nº 77 del Paseo de los Robles y los Martínez en el nº 79 carece de valor informativo para nuestros propósitos. Existe una relación definida entre la decisión de Guillermo de asistir al concierto y la dirección en que se mueven sus pies tras aparcar el coche. Hay una relación igualmente definida entre cierta actividad en las vías extrapiramidales del gobernador Morris por una parte y su acción de tachar «Nosotros los Estados» y escribir en su lugar «Nosotros el Pueblo» en el preámbulo de la Constitución de Estados Unidos por otra. Sin embargo, ¿no resultaría grotesco decir que, al dar cuenta de la individualización de los derechos de que disfrutaron los pri-

meros ciudadanos de Estados Unidos, hay que considerar dos niveles de explicación, y que uno de ellos remite a las vías extrapiramidales del gobernador Morris? Me parece un despropósito.

¿Y qué diremos de Kasparov y el ordenador que supuestamente «juega al ajedrez»? Llegó un momento en que el frustrado oponente de Deep Blue declaró que su adversario simplemente no estaba *jugando* al ajedrez. Carecía de pasión, era inmune a las presiones, no se enfrentaba a su adversario cara a cara. Recordemos las *Cartas sobre la educación estética del hombre* de Schiller, donde se nos dice que el hombre nunca es tan auténticamente él mismo como cuando juega. Pensemos en los amplios y diversos factores culturales y las disposiciones a la conducta que hay que reclutar para poder calificar de «juego» una determinada actividad, y luego cotejémoslos con cualquiera que sea el «proceso» que hace que Deep Blue mueva el alfil a b2. Deep Blue sólo «juega» al ajedrez en el mismo sentido en que el microondas «cocina» sopa, aunque la programación sea muchísimo más complicada.

¿Es posible que nos sintamos tentados a decir que, si ésta es la caracterización correcta, entonces también Kasparov juega al ajedrez en el mismo sentido en que el microondas cocina sopa, aunque la programación sea muchísimo más complicada? Esto es lo que, en última instancia, hace que la tesis de la IA fuerte resulte tan interesante. Si se adopta la «actitud intencional» de Dennett, no sólo es permisible, sino conceptualmente ventajoso, reconocer que Deep Blue posee cualesquiera que sean los motivos, sentimientos, creencias y actitudes que consideremos adecuado invocar en la explicación de lo que está haciendo Kasparov. De esta forma, Kasparov no es «reducido» al nivel de una máquina, sino que la máquina es elevada al rango de los sistemas inteligentes. Si tanto Kasparov como Deep Blue están ocultos detrás de una pantalla, y las preguntas relevantes del test de Turing reciben las mismas respuestas por parte de

ambos, entonces uno y otro son «inteligentes» en lo que respecta al nivel de preguntas respondidas con acierto. Pero entonces entramos en la Habitación China de Searle y empezamos a pensar que Deep Blue no es más que un dispositivo para clasificar fichas y que sus «respuestas» no son en propiedad respuestas sino simplemente «outputs». El debate se prolonga aún más, pero debido únicamente a una forma de histeria intelectual que hace que personas de una formación exquisita sean incapaces de reconocer lo estúpido de la proposición original, a saber, que Deep Blue *juega al ajedrez.*

Dennett llama la atención sobre la preferencia de Hacker por «sentido» y «sinsentido» como características relevantes de un argumento filosófico, en contraste con las de «verdadero» o «falso» de la ciencia. Las clasificaciones de Hacker, de una sobriedad tal vez excesiva, resultan trajes demasiado ceñidos para llevarlos a diario. Sea lo que fuere lo que se confecciona en los incansables talleres de la ciencia, el propio hecho de que sea objeto de posteriores correcciones, revisiones y ajustes deja claro que, para empezar, no se trataba de la «verdad»; tampoco era un sinsentido, o al menos un completo sinsentido, excepto en unos pocos casos palmarios. (Los objetos calentados no suben debido a que adquieran la sustancia de la *ligereza,* y el ojo seco no es signo de brujería.) El sinsentido, además, es una condena demasiado severa para posiciones filosóficas obtusas, excesivamente ambiciosas, gratuitamente autobiográficas o demasiado graves. (En mis tiempos de universitario, pensaba que el intento de Hume de reducir el concepto de causalidad a la concurrencia constante de ciertos objetos en la experiencia era una broma típicamente escocesa. No fue hasta más tarde que me vi forzado a aceptar la decepcionante conclusión de que ¡lo decía en serio!) Pero una crítica correcta de las clasificaciones de Hacker no se puede basar sólo en la irritación por su elección de palabras. Los sabios del Círculo de

Viena, afanados miembros de la Sociedad Ernst Mach, se inclinaban por considerar toda afirmación no empírica como un sinsentido literal. Los ingeniosos miembros del club Scriblerus —Pope, Swift, Arbuthnot—, tras leer las páginas de Locke sobre la identidad personal, concluyeron que éste había perdido los estribos filosóficos. Convengamos en que hay que empeñarse en que los tratados filosóficos sobre temas importantes sean claros, accesibles, coherentes y, lamentablemente, respetuosos con los temas escogidos. Quienes fracasan en tal empeño ofenden al sentido común. No es que el sentido común sea el árbitro definitivo; sólo que es el árbitro al que hay que ganarse si se quiere que el tratado ejerza alguna influencia más allá de las cuatro paredes del seminario. Si Spassky y Kasparov dudan de que un ordenador «juegue» al ajedrez, ¿no es Dennett quien debería reconsiderar el tema?

En este mismo sentido, Dennett afirma desenmascarar a Ryle y a Wittgenstein y demostrar que (como Hacker) son poco sinceros cuando sugieren que existen «reglas» que rigen el discurso filosófico, y no digamos el corriente. Califica de farol el *error categorial* de Ryle y su promesa de una «lógica de la existencia». Pese a todo el trabajo dedicado a la sintaxis, los lingüistas siguen devanándose los sesos con «El gato trepó árbol abajo». Así sea. Tampoco se puede proscribir «El desayuno fue un deleite para el hipotálamo, a la vista de cómo su conducta eléctrica se iba saciando a medida que avanzaba la comida». Lo que aquí se contraviene no es una ley, sino una convención —es decir, una *regla* cuyo incumplimiento, a diferencia de una ley, no se traduce en una comparecencia ante los tribunales sino en un malentendido—. Cuando las expresiones de este tipo se hacen habituales, los malentendidos se convierten en sistemáticos, preñados de paradojas no intencionadas e implicaciones imprevistas, suavizadas sólo en ocasiones por un involuntario aire humorístico.

Evidentemente, las «convenciones» del Politburo son diferentes de las del Parlamento Británico. Cuando hay que fijar con certeza el significado mismo, lo que importa es precisamente qué convenciones deben imperar. Esto plantea el dilema de decidir cuánta obediencia hay que rendir a la psicología popular. Dennett advierte a los neurocientíficos que actúen con «la máxima cautela» al manejar los términos de esta psicología, ya que, como él dice, «las presuposiciones de uso pueden subvertir sus propósitos». ¿Cómo? Convirtiendo «lo que de otro modo serían teorías y modelos empíricos prometedores en sinsentidos apenas disimulados». Si entiendo bien lo que aquí significa «presuposiciones de uso», diría que la presuposición central es que lo que Bennett y Hacker llaman «descripción psicológica corriente» debe hacer posible toda la variedad de interacciones prácticas y significativas entre los hablantes competentes. Es obvio que si su jerga vulgar no es más que un sinsentido apenas disimulado, cabe dudar de que posean los recursos lingüísticos para beneficiarse siquiera de la ilustración filosófica. Tal vez llegue el día en que el ciudadano de a pie convenga en hablar en términos de su «sistema encefálico triúnico», abandonando mansamente el viejo sinsentido apenas disimulado de las manzanas rojas, el césped verde y el cielo azul. Sin embargo, y por si acaso el género humano hubiera nacido en este juego de lenguaje, uno se pregunta de qué forma los neurocientíficos harían coincidir cualquier cosa en el «sistema encefálico triúnico» con —en efecto— el mundo tal como realmente se *ve*. Si el discurso popular pudiera en verdad corromper las teorías y los modelos empíricos, lo haría principalmente porque las teorías y los modelos no tienen una relación especial con este mismo discurso que, en definitiva, es el discurso de la vida real. Tal vez esto ayude a explicar por qué las teorías y los modelos que hoy se ofertan no son más que modelos de datos, eficaces sumarios de observacio-

nes abiertamente antisépticas suavizadas mediante manipulaciones supuestamente estadísticas, y presentadas como modelos altamente integrados de... *nadie*, ni siquiera de un cerebro.

Para defenderse de la acusación de incurrir él mismo en la falacia mereológica, Dennett cita sus propios escritos anteriores en los que establece una distinción entre los niveles personal y subpersonal de explicación. Habla de sí mismo como de un pionero en este sentido. Personalmente yo hubiera extendido los laureles a Aristóteles, quien nos recuerda que para explicar la ira, pongamos por caso, se puede hablar de cambios en la temperatura de la sangre o, en cambio, de la reacción al hecho de haber sido desairado:

> Un físico definiría una afección del alma de modo distinto a como lo haría un dialéctico; este último definiría, por ejemplo, la ira como el anhelo por devolver dolor por dolor [...] mientras que el primero la definiría como una ebullición de la sangre o de la sustancia caliente que rodea el corazón.[2]

De hecho, podríamos atribuir a Aristóteles las primeras reservas ante la falacia, ya que, en el mismo tratado, dice:

> Afirmar que es el alma la que está airada es tan inexacto como lo sería decir que es el alma la que teje telarañas o construye casas. Es indudablemente mejor [...] decir que es el hombre quien hace todo esto con su alma.[3]

La distinción más amplia entre las explicaciones que se basan en las causas y las que lo hacen en las razones es, como es bien sabido, venerable y no menos polémica por el hecho de que se la reconozca ampliamente como filosóficamente rigurosa. La falacia mereológica,

sin embargo, se manifiesta de diferentes modos. Es transparente cuando uno sugiere que la tarta de cumpleaños fue cortada por nervios eferentes de los niveles cervicales 5-8 de la médula espinal. No obstante, también actúa cuando se analiza la razón de la acción de Ronaldo como la suma de un gran número de mini razones. Así, la explicación correcta de que Ronaldo comprara el Toyota Prius sería que quería una ventaja mecánica mayor que la que se consigue al andar. Que Dennett es víctima de esta versión de la falacia queda patente con encomiable claridad en sus propias palabras:

> No atribuimos a las partes del cerebro una creencia [...] *cabal*; esto sí sería una falacia. No. Atribuimos a estas partes una especie atenuada de creencia.

El ejemplo propuesto es el del niño que «algo así como» cree que su papá es médico. Esto resulta poco convincente. Puede haber alguna duda sobre el objeto de la creencia, pero indudablemente no existe una «parte» de una creencia. Sin embargo, el propio ejemplo debe finalmente postular partes de creencias, pues «atenuar» una creencia mediante la acción de un sistema físico (por ejemplo, las partes del cerebro) es cambiar su valor a lo largo de una especie de continuo físico, y esto, en última instancia, es trabajar sobre sus «partes». Si decir que el cerebro tiene creencias parece una proclama de la Reina Roja de *Alicia en el país de las maravillas*, más propio de ella si cabe es afirmar que una parte del cerebro tiene creencias atenuadas.

Dennett es especialmente contundente cuando trata de refutar argumentos a favor de que el cerebro forma imágenes de uno u otro tipo. Declara acertadamente que la cuestión de si las estructuras del interior del cerebro están o no organizadas de modo que *funcionan* como

creadoras de imágenes es una cuestión empírica, más allá del ámbito de los modos filosóficos de análisis. Exactamente cómo reaccionan ante el mundo exterior constelaciones de neuronas interconectadas es una cuestión central en las ciencias del cerebro, y se ha abordado de forma especialmente provechosa en las investigaciones sobre el sistema visual. Nadie piensa de verdad que el mundo visual se proyecte sobre el cerebro como una *imagen* —y por supuesto no existen imágenes de olores o sonidos—. Más bien se buscan relaciones isomórficas entre las características ópticas del mundo visible y los patrones neuroeléctricos asociados con su inspección. Concedido todo esto, nos vemos enfrentados a una cuestión muy distinta: no la de cómo cierto algoritmo neuroeléctrico trata o «codifica» las propiedades ópticas del mundo visible, sino la de la relación entre esa codificación y lo que el perceptor afirma que (visualmente) ocurre. Considerar *esta* pregunta adecuada para guiar la investigación empírica es olvidarse de la propia esencia de la pregunta, ya que nada observado en el nivel de la neuroanatomía funcional «ve» en *ningún* sentido, ni siquiera en un sentido atenuado.

Menos tiempo se requiere para considerar la defensa que Dennett hace de LeDoux y otros autores de la *nomenklatura* de la neurociencia cognitiva. Decir que el «cerebro» puede saber de un peligro antes de que él o nosotros sepamos de qué se trata no puede ser más que una corrupción del lenguaje y, en el nivel de la explicación científica, una lamentable muesca en la navaja de Occam. El macaco neonato tiene células en la corteza auditiva que reaccionan a los gritos de angustia de esta especie. Estas células o la corteza no «saben» nada, al igual que un circuito de resistencia-capacitancia no «sabe» que se ha instalado una nevera, pese a que reacciona con una caída del voltaje. En la formación de criaturas que se enfrentarán a un mundo peligroso hay mucho de precableado y algo de cableado permanen-

te. La naturaleza las dota de la capacidad de hacer por instinto o de modo reflexivo lo que no puede esperar hasta completar los estudios superiores. Éste es el equipamiento que sortea *todo* aprendizaje y, por consiguiente, todo conocimiento. Dennett insiste en encajonar cualquier cantidad y variedad de hechos en un contenedor conceptual demasiado elástico para que tenga forma, y demasiado endeble para sostener el peso de los auténticos problemas. Su defensa consiste en señalar todos los pequeños hechos que se pueden embutir en ese contenedor pero, por lo general, la simple repetición de frases que suscitaron críticas en primer lugar se considera retóricamente ineficaz. Él y muchos otros del *movimiento* de la neurociencia cognitiva (porque tiene todas las características de un movimiento) han adoptado un idioma extraño para oídos que no sean los suyos propios, pero obtienen seguridad de la aparición de estas mismas frases en todos sus libros y artículos. La justificación que Dennett hace de su elección de locuciones chocantes basándose en la frecuencia con que él mismo las usa es —para adaptar un ejemplo de Wittgenstein— «como si alguien va a comprar varios ejemplares del periódico de la mañana para asegurarse de que lo que en él se dice es verdad».[4] ¿Y qué hay de Bennett y Hacker, en especial de este último? Hice una crítica muy favorable de su libro en *Philosophy*, y nada en las réplicas de John Searle o Daniel Dennett me haría reconsiderar mi juicio inicial. Consideré allí que los objetivos de los autores eran precisamente aquellos que constituyen la auténtica misión de la filosofía, la cual, en su proyección mayor, es nada menos que la crítica de la vida y, en sus ambiciones más modestas, una indagación crítica sobre nuestras afirmaciones epistémicas nucleares. La historia deja muy claras las consecuencias de cambiar esta misión por una posición más elevada en esa jerarquía en que la propia ciencia se sitúa. Hay algo muy importante en el modo en que Sófocles, a través de *Antígona*,

defiende los fundamentos morales de toda ley contra las pretensiones del rey. Pero Sófocles no realizó el trabajo al que se aplicaron Aristóteles, Cicerón, Tomás de Aquino y otros miembros de la tradición de la Ley Natural. Los hipocráticos acabaron sabiamente con la idea de una enfermedad divina, primero reconociendo cortésmente que los dioses lo ocasionan todo, y luego tratando cada enfermedad como una afección ni más ni menos divina que cualquier otra. Sin recomendar las camisas de fuerza conceptuales, yo propondría un régimen de contención y atención cuando personas serias (sean filósofos, científicos o simplemente partes interesadas) consideran el alcance y la autoridad de las reflexiones filosóficas y científicas sobre el *Lebenswelt*, el mundo de la vida. Uno puede y debe admirar *Antígona* y aprender de ella, sin exigir a continuación su inclusión en los códigos legales para dirimir disputas en los tribunales. Uno puede y debe admirar la protección que los hipocráticos se empeñaron en dotar a la medicina contra las convicciones religiosas rituales, sin insistir en que se prohíba rezar en las salas de urgencias. Y uno puede y debe admirar la investigación detallada, repetible y rigurosa de la comunidad neurocientífica, sin desarrollar una actitud escéptica hacia el sentido común y hacia el reconocimiento, por otro lado insistente, de nosotros mismos como nosotros mismos.

Tanto si el análisis que Peter Hacker hace de los despropósitos conceptuales de los neurocientíficos de hoy es enteramente sólido como si juicios posteriores lo condenan, no hay duda de que exhibe fidelidad a la misión filosófica a la vez que dominio de los recursos que la filosofía ha desarrollado para esa misión. No busca ser miembro honorario de la Sociedad de la Neurociencia, ni pretende enriquecer esa ya de por sí impresionante base de datos sobre la que se edifica el progreso científico. De hecho, los descubri-

mientos realmente importantes de la neurociencia cognitiva han sido obra de una reducida legión de especialistas, en su mayor parte desconocidos para los lectores del *Times Literary Supplement* y la *New York Review of Books*. Quienes situaron la ciencia de la visión en el mapa de la auténtica ciencia de primera fila no ofrecieron ninguna gran «neurofilosofía». Sus nombres no les dirán nada a la mayoría de los lectores de mi humilde escrito: Selig Hecht, M. H. Pirenne, Clarence Graham, H. K. Hartline y George Wald. Pitts y McCulloch permanecieron apegados a sus matemáticas y propusieron circuitos que, si se diseñan con inteligencia, arrojan unos resultados notables. Pavlov fue productivo cuando se aplicó a la química de la digestión, pero se convirtió en una especie de escritorzuelo cuando trató de traducir toda la psicología al lenguaje de la «cerebrodinámica». DuBois fue más sabio cuando se enfrentó a esta *quaestio vexata* y concluyó, estoy seguro que con alegría en su voz, *¡IGNORABIMUS!*

Si Hacker pretendiera entrar como socio en cualquier círculo de las ortodoxias, donde las cuotas se exigen en forma de clichés, lo haría en el del *anticartesianismo*. Al hablar de cliché no quiero decir nada más que una expresión o máxima manida. «Dios es bueno» es uno de estos clichés, y las personas creyentes tomarán la frecuencia de tal afirmación como medida de su verdad. «Rechazo el cartesianismo» es otra afirmación manida que también puede contener una reflexión sobre alguna verdad profunda o superior. Pero cartesianismo significa cosas distintas para los diferentes miembros del círculo anticartesiano. En un modo un tanto descuidado, se le culpa generalmente de una ontología de dos sustancias combinada con una teoría del «teatro-de-la-mente». A continuación, tanto el dualismo como el teatro interior se desestiman alegremente como pruebas de inocencia filosófica.

Presumo que todavía está permitido recordar a todos que Descartes fue el fundador de la geometría analítica, un auténtico maestro de la ciencia de la óptica, que comprendió en su totalidad la ciencia de su tiempo y que, en sus cartas y correspondencia, reunió a las mejores mentes de una época poblada por grandes pensadores. Pocas han sido las críticas que contra sus ideas se han hecho en algún momento de los dos últimos siglos que no hubieran anticipado Hobbes, Gassendi y el padre Mersenne, con los que Descartes mantuvo un animado debate en negro sobre blanco. Tal como le dijo a la princesa Elizabeth, le resultaba útil adoptar en sus escritos frases y analogías que quizá fueran demasiado filosóficas, para evitar que se le malinterpretara. Sabiendo lo que sabía sobre la materia, se sentía satisfecho de que el carácter esencial de la vida racional y perceptiva no se pudiera derivar de la materia en ninguna combinación posible. ¿Hay algo realmente ridículo en una ontología dualista que contrasta los entes extendidos y los no extendidos? Yo pienso que no; de hecho, *nadie piensa que sí*, porque queda descartado por el propio pensamiento.

¿Es esto un argumento a favor del dualismo? En distintos momentos a lo largo de un período quizá de cincuenta años, he reflexionado sobre cuántos tipos diferentes de «materia» podrían ser los constitutivos de la realidad en su totalidad. He logrado comprender dos tipos, que, a falta de mejores palabras, los llamaré tipo físico de materia y sea lo que fuere que sustenta las dimensiones moral, estética, racional y emocional de mi vida. Oh, llamémoslo «mental». En realidad, sólo es visible para mí una porción muy reducida de todo el espectro electromagnético; únicamente la «materia» que tiene una longitud de onda de entre 3.600 y 7.600 angstroms. Por si acaso mi capacidad de comprender la totalidad de la realidad es tan limitada como mi visión, lo mejor es dejar sin respuesta las preguntas sobre cuántos tipos distinguibles de entes componen «todo lo que hay». No

sé la cantidad. Dennett no la sabe. Hacker no la sabe. Pero ¿y si «la física está completa»? ¿Zanja esto la cuestión? Como un antiguo presidente de Estados Unidos parece que dijo: «Todo depende de lo que se entienda por "completo"».

Hacker pertenece a la escuela wittgensteniana de filosofía cuyos principales defensores y críticos son reclutados entre los filósofos mejor instruidos. La cuestión de si uno considera los temas filosóficos como «puzzles» que hay que encajar, o como «problemas» que hay que resolver, es un tema de gran calado. En cualquiera de ambas concepciones, son esenciales la claridad y la coherencia de la expresión. Nadie podría decir en serio que por no ser wittgensteiniano no está obligado a analizar el aparato cultural y lingüístico por el que se crean y comparten los conceptos.

Hacker escribe con precisión, hasta el punto de rayar en el pariente etimológico de ésta, la *preciosidad*. Cuida el detalle. Es posible que los lectores reaccionen ante tal cuidado de la forma en que reaccionamos ante los conductores que nunca exceden la velocidad permitida. Cuando dice que

> las verdades conceptuales perfilan el espacio lógico en el que se sitúan los hechos. Determinan qué tiene sentido.

se podría interpretar que está infravalorando los hechos u otorgando a la filosofía el poder de regirlos. No hace tal cosa. El cosmos está repleto de hechos, cuya inmensa pluralidad está más allá de nuestros sentidos e incluso de nuestra comprensión. De esta hoguera viva y radiante, extraemos unos pocos fragmentos —los visibles o casi visibles— y empezamos a tejer una historia. En raras ocasiones, la historia es tan sistemática, tan ajustada a los fragmentos de que disponemos que otras historias fluyen de la primera, y lue-

go otras, y pronto nos vemos poseídos por unos poderes virtualmente proféticos sobre cuáles serán las que vendrán a continuación. Es el filósofo, sin embargo, quien debe poner freno al entusiasmo de quienes cuentan las historias, ya que, abandonados a sus propios medios, podrían invocar un futuro que únicamente valide nuestra actual confusión.

NOTAS

Introducción a *Philosophical Foundations of Neuroscience*

Lo que sigue es el texto inalterado del prefacio de *Philosophical Foundations of Neuroscience* (en adelante, PFN), salvo por la reducción de los dos últimos párrafos y la eliminación de referencias cruzadas, sustituidas por notas cuando ha sido necesario.

1. Las objeciones metodológicas a estas distinciones se analizan a continuación y, con mayor detalle, en PFN, capítulo 14.

2. El capítulo 1 de PFN empieza, por consiguiente, con un estudio histórico de los primeros avances de la neurociencia.

3. De ahí que el capítulo 2 de PFN se dedique a un escrutinio crítico de sus compromisos conceptuales.

4. PFN §3.10.

5. Véase más adelante, en el extracto del capítulo 3 de PFN. El capítulo original es mucho más extenso que el extracto que aquí se expone y la argumentación es por tanto más detallada.

6. El reduccionismo se examina en el capítulo 13 de PFN.

7. En el capítulo 14 de PFN.

8. Véase PFN §14.3.

9. Véanse los capítulos 1 y 2 de PFN.

10. En PFN §6.31, donde se estudia la imaginería mental, y en PFN §8.2, donde se investiga el movimiento voluntario, se analizan ejemplos que demuestran la futilidad de la investigación.

11. En la discusión sobre la memoria en PFN §§5.21-5.22, y sobre las emociones y apetitos en PFN §7.1, se proporcionan ejemplos.

12. Nos ocupamos en detalle de las dificultades metodológicas en el capítulo 3, §3 de PFN (en este libro), y en el capítulo 14 de PFN.

Fragmento del capítulo 3 de *Philosophical Foundations of Neuroscience*

En estas páginas se reproduce el texto inalterado de las págs. 68-80 de PFN, a excepción de las referencias cruzadas, que, cuando ha sido necesario, se han relegado a las notas.

1. F. Crick, *The Astonishing Hipothesis* (Londres, Touchstone Books, 1995), págs. 30, 32 y sigs., 57.

2. G. Edelman, *Bright Air, Brilliant Fire* (Londres, Penguin Books, 1994), págs. 109 y sigs., 130.

3. C. Blakemore, *Mechanics of the Mind* (Cambridge, Cambridge University Press, 1977), pág. 91.

4. J. Z. Young, *Programs of the Brain* (Oxford, Oxford University Press, 1978), pág. 119 (trad. cast., *Los programas del cerebro*, México, Fondo de Cultura Económica, 1986).

5. A. Damasio, *Descartes' Error—Emotion, Reason and the Human Brain* (Londres, Papermac, 1996), pág. 173 (trad. cast. *El error de Descartes. La emoción, la razón y el cerebro humano*, Barcelona, Crítica, 1999).

6. B. Libet, «Unconscious cerebral initiative and the role of conscious will in voluntary action», *The Behavioural and Brain Sciences* (1985), 8, pág. 536.

7. J. P. Frisby, *Seeing: Illusion, Brain and Mind* (Oxford, Oxford University Press, 1980), págs. 8 y sigs. Resulta curioso en este punto que el

equívoco modo de hablar filosófico asociado con las tradiciones cartesiana y empirista, concretamente hablar del mundo «externo», se haya transferido de la mente al cerebro. Era equívoco porque pretendía contrastar un «mundo de la conciencia» interno con un «mundo de la materia» externo. Pero esto es confuso. La mente no es una especie de lugar, y lo que en este modo de hablar se dice que está *en* la mente no por ello está localizado espacialmente. De ahí también que el mundo (que no es «mera materia», sino también seres vivos) no está *espacialmente* «fuera de» la mente. El contraste entre lo que hay en el cerebro y lo que está fuera de él es, por supuesto, perfectamente literal e incuestionable. Lo cuestionable es la afirmación de que en el cerebro hay «descripciones simbólicas».

8. R. L Gregory, «The Confounded Eye», en R. L. Gregory y E. H. Gombric (comps.), *Illusion in Nature and Art* (Londres, Duckworth, 1973), pág. 50.

9. D. Marr, *Vision, a Computational Investigation into the Human Representation and Processing of Visual Information* (San Francisco, Freeman, 1980), pág. 3.

10. P. N. Johnson-Laird, «How could consciousness arise from the computations of the brain?», en C. Blakemore y S. Greenfield (comps.) *Mindwaves* (Oxford, Blackwell, 1987), pág. 257.

11. Susan Greenfield, al explicar a su audiencia los logros de la tomografía por emisión de positrones, anuncia admirada que por primera vez es posible *ver los pensamientos*. Semir Zeki informa a los miembros de la Royal Society que el nuevo milenio pertenece a la neurobiología, que, entre otras cosas, resolverá los antiquísimos problemas de la filosofía (véase S. Zeki, «Splendours and miseries of the brain», *Phil. Trans. R. Soc. Lond.* B (1999), 354, 2054. Véase PFN §14.42.

12. L. Wittgenstein, *Philosophical Investigations* (Oxford, Blackwell, 1953), §281 (trad. cast.: *Investigaciones filosóficas*, Barcelona, Crítica, 1988) (véanse también §§282-284, 357-361). La reflexión fundamental sobre esta observación la hizo A. J. P. Kenny, «The Homunculus Fallacy» (1971), reimpr. en su *The Legacy of Wittgenstein* (Oxford, Blackwell, 1984), págs. 126-136. Sobre la interpretación detallada de la observación de Witt-

genstein, véase P. M. S. Hacker, *Wittgenstein: Meaning and Mind, Volume 3 of an Analitical Comentary on the Philosophical Investigations* (Oxford, Blackwell, 1990), Exegesis §§281-184, 357-361, y el ensayo titulado «Men, Minds and Machines», que estudia algunas de las ramificaciones de la idea de Wittgenstein. Como bien se demuestra en el capítulo 1 de PFN, Aristóteles se le anticipó en esta idea (DA 408^{b}2-15).

13. Kenny (*ibid.*, pág. 125) emplea la expresión «falacia del homúnculo» para referirse al error conceptual en cuestión. Aunque sea una expresión muy gráfica, el autor admite que puede resultar confusa, ya que el error *no* es simplemente el de adscribir predicados psicológicos a un homúnculo imaginario que reside en la cabeza. En nuestra opinión, la expresión «falacia mereológica» es más adecuada. Hay que señalar, sin embargo, que el error en cuestión no es meramente la falacia de atribuir a una parte unos predicados que únicamente se aplican a un todo, sino que es un caso especial de esta confusión más general. Como señala Kenny, la errónea aplicación de un predicado no es, hablando estrictamente, una falacia, ya que no es una forma de razonamiento inválido, pero lleva a falacias (*ibid.*, págs. 125 y sigs.) No hay duda de que esta confusión mereológica es habitual tanto entre los psicólogos como entre los neurocientíficos.

14. Principios mereológicos comparables se aplican a objetos inanimados y a algunas de sus propiedades. Del hecho de que un coche es rápido no se sigue que su carburador sea rápido, y del hecho de que un reloj marque la hora con precisión no se sigue que su rueda volante señale la hora con exactitud.

15. Señalemos, no obstante, que cuando me duele la mano, soy yo quien sufre dolor, no mi mano. Y cuando alguien me hace daño en la mano, me hace daño a mí. Los verbos de sensación (a diferencia de los de percepción) se aplican a las partes del cuerpo, es decir, nuestro cuerpo es sensible y sus partes pueden doler, picar, etc. Pero las correspondientes expresiones verbales que incorporan sintagmas nominales, por ejemplo «tener dolor (picor, etc.)», sólo son predicables de la persona, no de sus partes (en las que se ubica la sensación).

16. Véase Simon Ullman, «Tacit Assumptions in the Computational Study of Vision», en A. Gorea (comp.), *Representations of Vision, Trends and Tacit Assumptions in Vision Research* (Cambridge, Cambridge University Press, 1991). El autor limita su exposición al uso (o, en nuestra opinión, el mal uso) de términos como «representación» y «representación simbólica».

17. La expresión es de Richard Gregory; véase «The Confounded Eye» en R. L. Gregory y E. H. Gombrich (comps.), *Illusion in Nature and Art* (Londres, Duckworth, 1973), pág. 51.

18. Véase C. Blakemore, «Understanding Images in the Brain», en H. Barlow, C. Blakemore y M. Wetson-Smith (comps.), *Images and Understanding* (Cambridge, Cambridge University Press, 1990), págs. 257-283 (trad. cast.: *Imagen y conocimiento*, Madrid, Crítica, 1994).

19. S. Zeki, «Abstraction and Idealism», *Nature,* 404 (abril de 2000), pág. 547.

20. J. Z. Young, *Programs of the Brain* (Oxford, Oxford University Press, 1978), pág. 192 (trad. cast., *Los programas del cerebro*, México, Fondo de Cultura Económica, 1986).

21. Brenda Milner, Larry Squire y Eric Kandel, «Cognitive Neuroscience and the Study of Memory», *Neuron,* 20 (1998), pág. 450.

22. Para una exposición detallada de esta cuestionable afirmación, véase PFN §5.22.

23. Ullmann, *ibid.*, págs. 314 y sigs.

24. Marr, *ibid.*, pág. 20.

25. Marr, *ibid.*, pág. 21.

26. Marr, *ibid.*

27. Para más críticas de la exposición computacional que Marr hace de la visión, véase PFN §4.24.

28. Frisby, *ibid.*, pág. 8.

29. Roger Sperry, «Lateral Specialization in the Surgically Separated Hemispheres», en F. O. Schmitt y F. G. Worden (comps.), *The Neurosciencies Third Study Programme* (MIT Press, Cambridge, Mass., 1974), pág. II. Para un análisis detallado de estas formas de descripción, véase PFN §14.3.

30. Blakemore, «Understanding Images in the Brian», pág. 265. Hay que señalar que lo que se necesita para reconocer el orden del cerebro *no* es un conjunto de *reglas*, sino simplemente un conjunto de correlaciones regulares. Una regla, a diferencia de la mera regularidad, es un criterio de conducta, una norma de corrección con la que se puede juzgar si la conducta es correcta o incorrecta, adecuada o inadecuada.

31. J. Z. Young, *Programs of the Brain* (Oxford University Press, Oxford, 1978), pág. 52. (trad. cast., *Los programas del cerebro*, México, Fondo de Cultura Económica, 1986).

32. Blakemore, *ibid.*, págs. 265-267.

33. J. Z. Young, *Programs of the Brain*, pág. II.

34. La confusión que genera la incapacidad de distinguir una regla de una regularidad, y lo normativo de lo causal, es evidente en los comentarios de Blakemore sobre el diagrama del «homúnculo» motor de Penfield y Rasmussen. Blakemore insiste en que «las mandíbulas y las manos están exageradamente representadas» («Understanding Images in the Brain», pág. 266, en la extensa nota aclaratoria sobre la Fig. 17.6); pero esto tendría sentido sólo si estuviéramos hablando de un mapa con un método de proyección que indujera a error —en este sentido, hablamos de las distorsiones relativas de la proyección (cilíndrica) de Mercator—. Pero, dado que todo lo que los dibujos representan es el número relativo de células causalmente responsables de determinadas funciones, *nada* está, o podría estar, exageradamente representado. Sin duda, Blakemore no quiere decir que en el cerebro haya más células correlacionadas causalmente con las mandíbulas y las manos de las que debería haber.

Fragmento del capítulo 10 de *Philosophical Foundations of Neuroscience*

1. Ned Block, «Qualia», en S. Guttenplan (comp.), *Blackwell Companion to the Philosophy of Mind* (Blackwell, Oxford, 1994), pág. 514.

2. R. Searle, «Consciousness», *Annual Review*, pág. 560.

3. Searle, *ibid.*, pág. 561.

4. Chalmers, *The Conscious Mind* (Oxford University Press, Oxford, 1996), pág. 4.

5. D. J. Chalmers, *The Conscious Mind*, pág. 10.

6. I. Glynn, *An Anatomy of Thought*, pág. 392.

7. A. Damasio, *The Feeling of What Happens*, pág. 9. Obsérvese que en este punto se da por supuesto, sin demostración alguna, que el color y el sonido no son propiedades de los objetos, sino de las impresiones sensoriales.

8. G. Edelman y G. Tononi, *Consciousness—How Matter Becomes Imagination*, pág. 157.

9. E. Lomand, «Consciousness», en *Routledge Encyclopaedia of Philosophy* (Routledge, Londres, 1998), vol. 2, pág. 581.

10. Searle, *The Mystery of Consciousness*, pág. XIV.

11. T. Nagel, «What it is like to be a bat?, reimpr. en *Mortal Questions* (Cambridge Universsity Press, Cambridge, 1979), pág. 166.

12. Nagel, *ibid.*, pág. 170n.

13. Nagel, *ibid.*, pág. 170.

14. M. Davies y G. W. Humphreys (comps.), *Consciousness* (Blackwell, Oxford, 1993), pág. 9.

15. Edelman y Tononi, *Consciousness—How Matter Becomes Imagination*, pág. 11.

16. Chalmers, *The Conscious Mind*, pág. 4.

17. Véase Searle, *The Mysteries of Consciousness*, pág. 201.

Fragmento del capítulo 14 de *Philosophical Foundations of Neuroscience*

1. Véase, por ejemplo, la discusión sobre los movimientos voluntarios en PFN §8.2.

2. Véase, por ejemplo, la discusión sobre la imaginería mental en PFN §14.3.

3. Véase, por ejemplo, PFN, §14.3.
4. Expuesto en PFN §4.23.
5. Tal como se dice en PFN §§6.3-6.31.
6. Véase PFN §2.3.

Neurociencia y Filosofía

1. Esta manera de usar términos como «almacenar» y «memoria» ha sido blanco de importantes críticas: véase PFN, págs. 158-171.

2. El profesor Dennett señala en su nota 15 que, durante la reunión anual de la APA , Bennett manifestó su «profunda consternación ante los modelos e hipótesis sensacionalistas de los neurocientíficos cognitivos actuales, y dejó claro que en su opinión eran todos ellos incomprensibles. Con un informante como Bennett, no es de extrañar que Hacker fuera incapaz de encontrar algo de valor en la neurociencia cognitiva». También sugiere que estoy irremisiblemente atrapado en esa «animadversión mutua» que impera entre los neurocientíficos sinápticos y los cognitivos. No es así. En primer lugar, David Marr es considerado un neurocientífico cognitivo de talento en los manuales sobre la materia (véase Gazzaniga, Ivry y Mangun, 2002, pág. 597); he publicado artículos sobre teoría de la red sináptica en el espíritu del trabajo de Marr, y no veo de forma alguna que ello demuestre una hostilidad irreflexiva hacia las neurociencias cognitivas (véase, por ejemplo, Bennett, Gibson y Robinson, 1994). Un libro de próxima publicación, obra de Hacker y mía, *History of Cognitive Neuroscience*, no se hubiera escrito si estuviéramos irremisiblemente atrapados en la hostilidad irracional ante la neurociencia cognitiva. En segundo lugar, en el encuentro de la APA no dije que «los modelos de los neurocientíficos cognitivos actuales» fueran «todos ellos incomprensibles». Lo que hice fue insistir en la extrema complejidad de la biología de la que son modelos y en la consecuente indigencia de nuestros conocimientos biológicos. Esto hace muy difícil construir modelos que esclarezcan las funciones de las redes

sinápticas. En el segundo apartado de este capítulo se proporcionan ejemplos invocados en apoyo de esta idea. Sin embargo, en la reunión de la APA añadí que resulta extraño decir de tales redes y grupos de redes que «ven», «recuerdan», etc., es decir, que poseen los atributos psicológicos de los seres humanos (véase el tercer apartado).

LA FILOSOFÍA COMO ANTROPOLOGÍA INGENUA

1. Mi objetivo en *Content and Consciousness* (trad. cast.: *Contenido y conciencia*, Barcelona, Gedisa, 1996) de 1969 era «establecer el marco conceptual desde el que hay que contar toda la historia, determinar las restricciones que se imponen a cualquier teoría satisfactoria (pág. ix) [...] [desarrollar] la idea de un modo distintivo de discurso, el lenguaje de la mente, que normalmente usamos para describir y explicar nuestras experiencias mentales, y que se puede relacionar sólo indirectamente con el modo de discurso en que se formula la ciencia» (pág. x).

2. Aunque teóricos anteriores —por ejemplo, Freud— hablaron de *psicología popular* en un sentido algo distinto, creo que yo fui el primero, en «Three Kinds of Intentional Psychology» (1978), en proponer su uso para denominar lo que Hacker y Bennett llaman «descripción psicológica corriente». Ellos, al igual que yo, insisten en que no se trata de una *teoría*.

3. Véase mi exposición sobre este punto en «A Cure for the Common Code», en *Brainstorms* (1978) y, más recientemente, en «Intentional Laws and Computacional Psychology» (apartado 5 de «Back from de Drawing Board») en Dahlbom (comp.), *Dennett and His Critics*, 1993.

4. La lista es larga. Véanse mis críticas, además de las obras citadas en las notas anteriores, a las obras sobre la imaginería, los *qualia*, la introspección y el dolor en *Brainstorms*. No soy el único teórico cuyo trabajo omiten citar pese a que se anticipa al suyo. Por ejemplo, en la explicación que hacen de la imaginería mental, reinventan sin darse cuenta varias ideas de Zenon Pylyshyn. Bennett y Hacker no son los primeros analis-

tas conceptuales en frecuentar estas aguas, y la mayor parte de sus puntos, si no todos, han sido anteriormente expresados y debidamente examinados en obras que no citan. No encontré nada nuevo en su libro.

5. El apéndice que dedican a atacar mis ideas es un extenso comentario despectivo, una colección de interpretaciones estúpidamente erróneas que acaba así: «Si nuestros argumentos son válidos, entonces las teorías de la intencionalidad y de la conciencia de Dennett no aportan nada al esclarecimiento filosófico de la intencionalidad o la conciencia. Tampoco ofrecen orientación alguna para la investigación neurocientífica o la comprensión neurocientífica» (pág. 435). Pero no hay argumentos, sólo declaraciones de «incoherencia». En la reunión anual de la APA en la que se expuso este ensayo, Hacker respondió con más de lo mismo. En el Oxford de la década de 1960, solía ocurrir que un delicado encogimiento de hombros se tenía por un argumento. Pasaron aquellos tiempos. Mi consejo a Hacker: si encuentra usted estos temas incomprensibles, pruebe estudiarlos con mayor esfuerzo. Apenas ha iniciado usted su formación en ciencia cognitiva.

6. Hornsby, 2000. El hecho de que Hacker haga caso omiso de mi distinción no se puede atribuir a la miopía; además de Hornsby, la han discutido en profundidad otros filósofos de Oxford, por ejemplo, Davies, 2000; Hurley, *Synthese*, 2001; y Bermúdez, «Nonconceptual Content: From Perceptual Experience to Subpersonal Computational States», *Mind and Language*, 1995.

7. Véase también «Conditions of Personhood», en *Brainstorms*.

8. Véase también la exposición de los niveles de explicación en *La conciencia explicada* (1991).

9. Searle, que presentó una letanía de objeciones a Bennett y Hacker, apenas mencionó esta cuestión en el encuentro de la APA en que se presentó este ensayo.

10. Para un filósofo que no toma la verdad y la falsedad como las piedras clave de las proposiciones filosóficas, Hacker es extremadamente liberal en sus crudas aseveraciones sin argumentar que tal cosa está mal interpretada, tal otra es un error y demás. Estos comentarios pasajeros son difíci-

les de interpretar sin la suposición de que se pretende que sean verdaderos (en oposición a falsos). Tal vez debamos entender que sólo una mínima fracción de las proposiciones de Hacker, aquellas específicamente filosóficas, «anteceden» la verdad y la falsedad, mientras la inmensa mayoría de sus frases son lo que parecen ser: afirmaciones que apuntan a la verdad. Y como tales, presumiblemente, están sometidas a la confirmación y la refutación empíricas.

11. En *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness* (2005) (trad. cast.: *Dulces sueños: obstáculos filosóficos para una ciencia de la conciencia*, Katz Barpal, 2006), describo algunas tensiones de la filosofía de la mente contemporánea como *autoantropología apriorística ingenua* (págs. 31-35). La obra de Hacker se me antoja un caso paradigmático.

12. Obsérvese que no digo que la autoantropología sea siempre una empresa insensata o inútil. Lo que digo es que es una indagación empírica que produce resultados —cuando se actúa correctamente— sobre las intuiciones que los investigadores descubren en ellos mismos, y las implicaciones de esas intuiciones. Estos resultados pueden ser fructíferos para la investigación, pero está por ver en qué condiciones pueden considerarse seriamente estas implicaciones como guías hacia la verdad en cualquier tema. Para más detalles al respecto, véase *Sweet Dreams*.

13. *The Claim of Reason: Wittgenstein, Skepticism, Morality, and Tragedy* (1979, 2ª ed., 1999) (trad. cast.: *Reivindicaciones de la razón*, Madrid, Síntesis, 2003).

14. ¿Es posible que un filósofo como Hacker *tenga razón* incluso cuando su objetivo no es la verdad?

15. Cabe suponer que Bennett, distinguido neurocientífico, ha representado el papel de informante para el antropólogo Hacker, pero entonces, ¿cómo explicar la casi completa insensibilidad de Hacker ante las sutilezas de la *jerga* (y los modelos y los descubrimientos) de la ciencia cognitiva? ¿Es que Hacker ha escogido el informante equivocado? Quizá las investigaciones de Bennett en la neurociencia se han realizado en el nivel

de la sinapsis, y quienes trabajan en este nivel subneuronal están tan lejos de las disciplinas de la ciencia cognitiva como los biólogos moleculares lo están de los etólogos de campo. No hay mucha comunicación entre tan distantes empresas e, incluso en las mejores circunstancias, existe mucha incomprensión —y una gran cantidad de animadversión mutua, por triste que sea decirlo—. Recuerdo a un distinguido director de laboratorio que iniciaba un taller con la siguiente observación: «En nuestro laboratorio tenemos un dicho: si se trabaja en una neurona, se trata de neurociencia; si se trabaja en dos neuronas, se trata de psicología». No lo decía como un cumplido. Escoger un informante hostil es, por supuesto, la receta del desastre antropológico. (Añadido después del encuentro de la APA.) Bennett confirmó esta conjetura en sus observaciones iniciales; después de repasar su carrera de estudioso de la sinapsis, manifestó su profunda consternación ante los modelos e hipótesis sensacionalistas de los neurocientíficos cognitivos actuales, y dejó claro que en su opinión eran todos ellos incomprensibles. Con un informante como Bennett, no es de extrañar que Hacker no supiera encontrar nada de valor en la neurociencia cognitiva.

16. Véase mi *Content and Consciousness*, pág. 183 (trad. cast.: *Contenido y conciencia*, Barcelona, Gedisa, 1996).

17. Por poner sólo un ejemplo, cuando Hacker deplora mi «bárbara locución verbal "ser cuestión de"» (pág. 422) e insiste en que «el receptor opioide *es tan cuestión* de opioides como los gatos lo son de perros, o los zorros de zorras» (pág. 423), está por supuesto completamente en lo cierto: la elegante relación entre los opioides y el receptor opioide no es «una cuestión de» cabal (lamento lo intrincado de los términos), sino una mera «protocuestión de» (¡y dale!), pero éste es *precisamente* el tipo de propiedad que uno podría apreciar en una mera parte de una determinada suma mereológica que (adecuadamente organizada) podría mostrar una intencionalidad de buena fe, auténtica, filosóficamente sólida, paradigmática...

18. En el sentido restringido de Hacker.

19. Éste ha sido un tema recurrente en las obras críticas de la ciencia cognitiva. Los artículos clásicos se remontan al de William Word «What's in a

Link» (en Bobrow y Colins, *Representation and Understanding*, 1975) y al de Drew McDermott «Artificial Intelligence Meets Natural Stupidity», en Haugeland, *Mind Design* (1981), pasando por el de Neisser, *Cognition and Reality* (1975) (trad. cast.: *Procesos cognitivos y realidad*, Madrid, Marova, 1981) y el de Rodney Brook «Intelligence Without Representation», *Artificial Intelligence* (1991). La lista sigue hasta hoy en día, incluidas las aportaciones de filósofos que han hecho los deberes y conocen al detalle estos temas.

20. Dicen Hacker y Bennett: «Sería un error, por otra parte inocuo, hablar de mapas del cerebro cuando lo que se quiere decir es que ciertas características del campo visual se pueden mapear sobre las descargas de grupos de células de la corteza visual primaria o estriada. Pero de ahí no se puede pasar a decir, como hace Young, que el cerebro hace uso de sus mapas cuando formula hipótesis sobre lo que es visible» (pág. 77). Pero esto es lo que hace que hablar de mapas sea perspicuo: que el cerebro *hace* uso de ellos *como* mapas. De no ser así, es evidente que este modo de hablar no tendría sentido. Y ésta es la razón de que el énfasis de Kosslyn en los patrones visibles de excitación en la corteza durante la imaginería sea completamente irrelevante en lo que se refiere a la naturaleza de los procesos subyacentes a lo que llamamos, en el nivel personal, imaginería visual. Véase el artículo central de Pylyshyn en el número de abril de 2002 de *BBS* y mi comentario, «*Does Your Brain Use the Images on It, and If So, How?*».

21. «Los filósofos no deberían hallarse en la situación de tener que abandonar sus teorías favoritas sobre la naturaleza de la conciencia ante la evidencia de los resultados científicos. No deberían tener teorías favoritas, pues en primer lugar no deberían estar postulando teorías empíricas sometidas a la confirmación y refutación empíricas. Lo suyo son los conceptos, no los juicios empíricos; las formas del pensamiento, no su contenido; lo que es lógicamente posible, no lo que es empíricamente real; lo que tiene y lo que no tiene sentido, no lo que es y no es verdad» (pág. 404). Esta estrechez de miras acerca del negocio propio del filósofo es lo que hace que Hacker yerre tan estrepitosamente cuando trata de criticar a los científicos.

22. Para un ejemplo de este *tipo* de explicación, véase la mía simplificada de cómo el robot Shakey distingue las cajas de las pirámides (un talento de «nivel personal» en un robot) trazando (en el nivel subpersonal) dibujos lineales de sus imágenes retinales y usando después su programa de semántica lineal para identificar las características distintivas de las cajas, en *La conciencia explicada.*

23. El «Apéndice I: Daniel Dennett» de Bennett y Hacker no merece una réplica detallada, considerando los frecuentes errores de lectura de pasajes citados fuera de contexto, y la omisión, al parecer voluntaria, de cualquier discusión de los pasajes donde me defiendo específicamente de los errores de lectura que repiten una y otra vez, como ya he señalado. Sin embargo, no me puedo resistir a señalar que caen engañados por el bulo creacionista que supone anticiparse a cualquier explicación de las características biológicas en términos de lo que yo denomino la actitud del diseño: «La evolución no ha *diseñado* nada —el logro de Darwin fue desplazar la explicación en términos de diseño por explicaciones evolutivas» (pág. 425). Al parecer no entienden cómo funciona la explicación evolutiva.

Situar de nuevo la conciencia en el cerebro

Estoy en deuda con Romelia Drager, Jennifer Hudin y Dagmar Searle, por sus comentarios a los primeros borradores de este artículo.

1. Por ejemplo, John R. Searle, *The Rediscovery of the Mind*, Cambridge, MIT Press, 1992 (trad. cast.: *El redescubrimiento de la mente*, Barcelona, Crítica, 1996).

2. Para posibles contraejemplos de esta afirmación, véase por ejemplo la explicación de Christof Koch sobre «la neurona de Halle Berry», *New York Times*, 5 de julio, 2005.

3. John R. Searle, *Rationality in Action* (Cambridge, MIT Press, 2001).

Los supuestos conceptuales de la neurociencia cognitiva

1. M. R. Bennett y P. M. S. Hacker, *Philosophical Foundations of Neuroscience* (Oxford, Blackwell, 2003); las referencias a este libro se harán mediante las iniciales PFN.

2. El profesor Searle afirma que un resultado conceptual es importante sólo como parte de una teoría general (pág. 122). Si por «una teoría general» entiende una explicación de conjunto de una red conceptual, y no meros resultados aislados, estamos de acuerdo con él. Cuando negamos que nuestras explicaciones generales sean teóricas, nos referimos a que no están en el mismo nivel lógico que las teorías científicas. Son descripciones, no hipótesis; no se pueden confirmar ni refutar mediante experimentos; no son hipotético-deductivas, y su objetivo no es ni la predicción ni proporcionar explicaciones causales; no implican idealizaciones en el sentido en que lo hacen las ciencias (por ejemplo, la idea de masa de un punto en la mecánica newtoniana), y no se aproximan a los hechos empíricos dentro de unos márgenes de error convenidos; no se pretenden descubrir nuevas entidades, ni se introducen entidades hipotéticas con fines explicativos.

3. Parece que el profesor Dennett tenía problemas con esta idea. En sus críticas (pág. 79), citaba de forma selectiva de nuestro libro: «Las preguntas conceptuales son previas a las cuestiones de verdad y falsedad...» (PFN, pág. 2, véase pág. 4 en este libro), «La verdad y la falsedad son a la ciencia lo que el sentido y el sinsentido son a la filosofía» (PFN, pág. 6, véase pág. 12 de este libro). A partir de ahí sacaba la conclusión de que, desde nuestra concepción, a la filosofía no le interesa en absoluto la verdad. Sin embargo, omitía la continuación de la primera frase:

> Son preguntas que conciernen a nuestras formas de representación, no a la verdad o la falsedad *de afirmaciones empíricas*. Estas formas están presupuestas en las afirmaciones científicas verdaderas (y en las falsas) y las teorías científicas correctas (e incorrectas). No determinan lo que es

empíricamente verdadero o falso, sino más bien lo que tiene o no tiene sentido.

(PFN, pág. 2, véase pág. 4 de este libro; la cursiva es nuestra)

Asimismo, el profesor Dennett omitía la observación de la página contigua de que la neurociencia está descubriendo muchas cosas relativas a los fundamentos neurales de las capacidades humanas, «pero sus descubrimientos no afectan en modo alguno a *la verdad conceptual* de que estas capacidades y su ejercicio [...] son atributos de los seres humanos, no de sus partes» (PFN, pág. 3, véase pág. 6 de este libro; la cursiva es nuestra). Como resulta patente, nuestra concepción es que la filosofía se ocupa de las verdades conceptuales, y que éstas determinan qué tiene sentido y qué no.

4. El profesor Paul Churchland dirige la siguiente observación contra nuestra concepción: «desde Quine, el grueso de la profesión filosófica se ha inclinado por decir "no"» a la sugerencia de que hay «verdades necesarias, constitutivas de significados, situadas para siempre más allá de toda refutación empírica o factual». «Cleansing Science», *Inquiry*, 48 (2005), pág. 474. Es dudoso que haya llevado a cabo una encuesta sociológica (¿es que la mayoría de los filósofos realmente piensan que las verdades de la aritmética están sometidas a la refutación empírica, junto a cualquier teoría empírica en la que se integren?), y nos sorprende que un filósofo pueda pensar que un recuento de cabezas es un criterio de verdad.

5. Para una crítica canónica de lo que Quine dice sobre la analiticidad, véase P. F. Strawson y H. P. Grice, «In Defense of a Dogma», *Philosophical Review*, 1956. Para una crítica más reciente y meticulosa de la posición global de Quine, véase H.-J. Glock, *Quine and Davidson on Language, Thought, and Reality* (Cambridge, Cambridge University Press, 2003). Sobre los contrastes entre Quine y Wittgenstein, véase P. M. S. Hacker, *Wittgenstein's Place in Twentieth-Century Analytic Philosophy* (Oxford, Blackwell, 1996), cap. 7.

6. Se podría pensar (como sugería el profesor Churchland) que la idea de Descartes de que la mente puede afectar causalmente al movimiento

del cuerpo (entendida, según el profesor Churchland, como una afirmación conceptual) queda refutada por la ley de la conservación del momento. Sería un error. La afirmación cartesiana (con independencia de que sea conceptual o empírica) únicamente podría refutarse si tuviera sentido; pero, en ausencia de un criterio de identidad para las sustancias inmateriales, no lo tiene. La idea misma de que la mente es una *sustancia* de algún tipo no es coherente. De ahí que la afirmación de que la mente, así entendida, posee poderes causales no sea inteligible, *a fortiori* ni confirmable ni refutable por la observación y la comprobación experimentales. (Pensemos en cuál sería el resultado experimental que *contaría* como demostración de que es verdadera.)

7. Tal concepción *epistémica* subyace al extenso ataque del profesor Timothy Williamson a la idea misma de verdad conceptual, «Conceptual Truth», *Proceedings of the Aristotelian Society*, supl. vol. 80 (2006). La concepción que esboza no es lo que muchos grandes pensadores, desde Kant hasta nuestros días, entienden por «una verdad conceptual». Después de criticar, para su propia satisfacción, la concepción epistémica que él mismo acaba de perfilar, el profesor Williamson saca la conclusión de que *no hay verdades conceptuales de ningún tipo*. Pero esta conclusión no se sigue en absoluto. Porque todo lo que ha demostrado (en el mejor de los casos) es que no existen verdades conceptuales que quepan en la cama de Procrusto epistémica que ha ideado.

8. Los puntos aristotélicos y anticartesianos que destacamos son: 1) el principio de Aristóteles, del que nos ocupamos más abajo; 2) la identificación aristotélica de la *psuchē* con una serie de capacidades; 3) que estas capacidades se identifican por lo que capacitan para hacer; 4) que para determinar si una criatura posee una capacidad hay que observar sus actividades; 5) la observación de Aristóteles de que preguntarse si la *psuchē* y el cuerpo son una cosa o dos es una cuestión incoherente.

9. No es, por supuesto, una falacia en sentido estricto, pero lleva a incurrir en falacias: inferencias inválidas y argumentos falsos.

10. A. J. P. Kenny, «The Homunculus Fallacy», en M. Grene (comp.),

empíricamente verdadero o falso, sino más bien lo que tiene o no tiene sentido.

(PFN, pág. 2, véase pág. 4 de este libro; la cursiva es nuestra)

Asimismo, el profesor Dennett omitía la observación de la página contigua de que la neurociencia está descubriendo muchas cosas relativas a los fundamentos neurales de las capacidades humanas, «pero sus descubrimientos no afectan en modo alguno a *la verdad conceptual* de que estas capacidades y su ejercicio [...] son atributos de los seres humanos, no de sus partes» (PFN, pág. 3, véase pág. 6 de este libro; la cursiva es nuestra). Como resulta patente, nuestra concepción es que la filosofía se ocupa de las verdades conceptuales, y que éstas determinan qué tiene sentido y qué no.

4. El profesor Paul Churchland dirige la siguiente observación contra nuestra concepción: «desde Quine, el grueso de la profesión filosófica se ha inclinado por decir "no"» a la sugerencia de que hay «verdades necesarias, constitutivas de significados, situadas para siempre más allá de toda refutación empírica o factual». «Cleansing Science», *Inquiry*, 48 (2005), pág. 474. Es dudoso que haya llevado a cabo una encuesta sociológica (¿es que la mayoría de los filósofos realmente piensan que las verdades de la aritmética están sometidas a la refutación empírica, junto a cualquier teoría empírica en la que se integren?), y nos sorprende que un filósofo pueda pensar que un recuento de cabezas es un criterio de verdad.

5. Para una crítica canónica de lo que Quine dice sobre la analiticidad, véase P. F. Strawson y H. P. Grice, «In Defense of a Dogma», *Philosophical Review*, 1956. Para una crítica más reciente y meticulosa de la posición global de Quine, véase H.-J. Glock, *Quine and Davidson on Language, Thought, and Reality* (Cambridge, Cambridge University Press, 2003). Sobre los contrastes entre Quine y Wittgenstein, véase P. M. S. Hacker, *Wittgenstein's Place in Twentieth-Century Analytic Philosophy* (Oxford, Blackwell, 1996), cap. 7.

6. Se podría pensar (como sugería el profesor Churchland) que la idea de Descartes de que la mente puede afectar causalmente al movimiento

del cuerpo (entendida, según el profesor Churchland, como una afirmación conceptual) queda refutada por la ley de la conservación del momento. Sería un error. La afirmación cartesiana (con independencia de que sea conceptual o empírica) únicamente podría refutarse si tuviera sentido; pero, en ausencia de un criterio de identidad para las sustancias inmateriales, no lo tiene. La idea misma de que la mente es una *sustancia* de algún tipo no es coherente. De ahí que la afirmación de que la mente, así entendida, posee poderes causales no sea inteligible, *a fortiori* ni confirmable ni refutable por la observación y la comprobación experimentales. (Pensemos en cuál sería el resultado experimental que *contaría* como demostración de que es verdadera.)

7. Tal concepción *epistémica* subyace al extenso ataque del profesor Timothy Williamson a la idea misma de verdad conceptual, «Conceptual Truth», *Proceedings of the Aristotelian Society*, supl. vol. 80 (2006). La concepción que esboza no es lo que muchos grandes pensadores, desde Kant hasta nuestros días, entienden por «una verdad conceptual». Después de criticar, para su propia satisfacción, la concepción epistémica que él mismo acaba de perfilar, el profesor Williamson saca la conclusión de que *no hay verdades conceptuales de ningún tipo*. Pero esta conclusión no se sigue en absoluto. Porque todo lo que ha demostrado (en el mejor de los casos) es que no existen verdades conceptuales que quepan en la cama de Procrusto epistémica que ha ideado.

8. Los puntos aristotélicos y anticartesianos que destacamos son: 1) el principio de Aristóteles, del que nos ocupamos más abajo; 2) la identificación aristotélica de la *psuchē* con una serie de capacidades; 3) que estas capacidades se identifican por lo que capacitan para hacer; 4) que para determinar si una criatura posee una capacidad hay que observar sus actividades; 5) la observación de Aristóteles de que preguntarse si la *psuchē* y el cuerpo son una cosa o dos es una cuestión incoherente.

9. No es, por supuesto, una falacia en sentido estricto, pero lleva a incurrir en falacias: inferencias inválidas y argumentos falsos.

10. A. J. P. Kenny, «The Homunculus Fallacy», en M. Grene (comp.),

Interpretations of Life and Mind (Londres, Routledge, 1971). Nosotros preferimos el nombre menos gráfico pero más preciso de «falacia mereológica» (y, correlativamente, «el principio mereológico»). Descubrimos que los neurocientíficos eran propensos a desechar, por infantil, la falacia de suponer que en el cerebro habita un homúnculo y de ahí pasar directamente a adscribir atributos psicológicos al cerebro.

11. Evidentemente no *con* su cerebro, en el sentido en que uno hace cosas *con* las manos o los ojos, ni en el sentido en que uno hace cosas con sus habilidades. Por cierto, no podría hacer ninguna de estas cosas si su cerebro no funcionara normalmente.

12. D. Dennett, *Content and Consciousness* (Londres, Routledge and Kegan Paul, 1969) (trad. cast.: *Contenido y conciencia*, Barcelona, Gedisa, 1996).

13. Nos sorprendió bastante ver que el profesor Dennett declaraba que sus «principales puntos de desacuerdo» son que no cree que «el nivel personal de explicación sea el *único* nivel de explicación cuando se trata de la mente y la acción humanas», y que piensa que la tarea de relacionar estos dos niveles de explicación «no está fuera del ámbito del filósofo» (pág. 79). No hay desacuerdo alguno en lo que a *esto* se refiere. Cualquiera que en algún momento se haya tomado una aspirina para aliviar el dolor de cabeza, o que haya tomado alcohol suficiente para sentirse alegre, pendenciero o taciturno, y quiere una explicación de la secuencia de acontecimientos, debe compartir sin duda la primera idea de Dennett. Cualquiera que, como nosotros en las 452 páginas de *Philosophical Foundations of Neuroscience*, se haya dedicado a esclarecer las relaciones lógicas entre los conceptos psicológicos y los neurológicos, y entre los fenómenos que expresan, compartirá la segunda idea.

14. L. Wittgenstein, *Philosophical Investigations* (Oxford, Blakwell, 1953), §281 (trad. cast.: *Investigaciones filosóficas*, Barcelona, Crítica, 1988).

15. La concepción cartesiana del cuerpo que tiene un ser humano está equivocada en muchos aspectos. Descartes concebía su cuerpo como una

máquina insensible —una sustancia material sin sensación—. Pero nuestra auténtica concepción del cuerpo predica verbos de sensación del cuerpo que tenemos —es nuestro cuerpo lo que nos duele por todas partes o nos pica tanto que no lo podemos resistir.

16. El cerebro humano es parte del ser humano. También se puede decir que es parte del cuerpo que se dice que un ser humano tiene. Sin embargo, resulta significativo que dudemos en decir de una persona viva, a diferencia de un cadáver, que su cuerpo *tiene* dos piernas o, de alguien a quien se le haya amputado un miembro, que su cuerpo *tiene* sólo una pierna. El engañoso posesivo se aplica al ser humano y un cadáver humano, pero no, o sólo de forma vacilante, al cuerpo que se dice que tiene un ser humano viviente. Aunque el cerebro es una parte del cuerpo humano, no hay duda de que no diríamos «mi cuerpo *tiene* un cerebro» o «el cerebro de mi cuerpo tiene meningitis». Y no es por casualidad.

17. Convenimos con el profesor Searle en que la pregunta de qué animales inferiores son conscientes no se puede resolver mediante el «análisis lingüístico» (pág. 104). Pero él supone que se podría resolver mediante la investigación de sus sistemas nerviosos, mientras que nosotros señalamos que se podría resolver con el estudio de la conducta que el animal muestra en las circunstancias de su vida. Del mismo modo que establecemos si un animal ve mediante la referencia a sus reacciones ante lo visible, también averiguamos si un animal es capaz de tener conciencia mediante el estudio de su repertorio conductual y sus reacciones ante su entorno. (Esto no implica que ser consciente sea comportarse de una determinada forma, sino sólo que los criterios para ser consciente son conductuales.)

18. La justificación de aplicar predicados psicológicos a los demás se sustenta en las pruebas disponibles. Éstas pueden ser de tipo inductivo o constitutivas (criterios). Las pruebas inductivas, en estos casos, presuponen razones no inductivas, criterios. Los criterios para la aplicación de un predicado psicológico están en la conducta (no meramente en los movimientos corporales) en las circunstancias adecuadas. Los criterios son corregibles. El hecho de que tales o cuales razones justifiquen la atribución de un pre-

dicado psicológico a alguien es en parte constitutivo del significado de predicado, pero no agota su significado. Los criterios para la aplicación de tal predicado son distintos de sus condiciones de verdad —un animal puede estar sintiendo un dolor y no demostrarlo o mostrar una conducta de dolor sin padecerlo— (no somos conductistas). Las condiciones de verdad de una proposición que atribuya un predicado psicológico a un ser son distintas de la verdad de la proposición. Tanto los criterios como las condiciones de verdad son distintos de las condiciones generales bajo las que se pueden llevar a cabo de forma significativa las actividades de aplicar o negar el predicado a criaturas. Pero es un error suponer que una condición del «juego del lenguaje que se está jugando» (en expresión del profesor Searle) es la *ocurrencia* de conducta públicamente observable. Y es que el juego del lenguaje con un predicado psicológico se juega con su negación no menos que con su afirmación. También sería un error mezclar las condiciones para el aprendizaje de un juego del lenguaje con las que se requieren para jugarlo.

19. J. Z. Young, *Programs of the Brain* (Oxford University Press, 1978), pág. 192 (trad. cast.: *Los programas del cerebro humano*, México, Fondo de Cultura Económica, 1986). El profesor Dennett también sugiere (pág. 90) que tergiversamos a Crick cuando sostenemos que, porque éste dice que nuestro cerebro cree cosas y hace representaciones sobre la base de su experiencia o información previas (F. Crick, *The Astonishing Hypoyhesis*, Londres, Touchstone, 1995, págs. 28-33, 57 [trad. cast.: *La búsqueda científica del alma. Una revolucionaria hipótesis para el siglo XXI*, Madrid, Debate, 1994]), Crick realmente piensa que el cerebro cree cosas y hace interpretaciones, etc. Invitamos al lector a que lea él mismo las citadas explicaciones de Crick.

20. N. Chomsky, *Rules and Representations*, Oxford, Blackwell, 1980 (trad.cast.: *Reglas y representaciones*, México, Fondo de Cultura Económica, 1983). Lejos de hacer caso omiso a esto, como afirmaba el profesor Dennett (pág. 91), el tema se trató con ojo crítico en G. P. Baker y P. M. S. Hacker, *Language, Sense and Nonsense*, Oxford, Blackwell, 1984, págs. 340-45.

21. D. Dennett, *Consciousness Explained* (Harmondsworth, Penguin, 1993), págs.142-144 (trad. cast.: *La conciencia explicada*, Barcelona, Paidós, 1995).

22. Dennett cita aquí su entrada autobiográfica en S. Guttenplan (comp.), *A Companion to the Philosophy of Mind* (Oxford, Blackwell, 1994), pág. 240.

23. Evidentemente, no negamos que la extensión analógica de los conceptos y las estructuras conceptuales a menudo es útil en la ciencia. La analogía hidrodinámica generó una teoría de la electricidad provechosa, comprobable y matematizada. Nada comparable es manifiesto en la licencia poética de la actitud intencional de Dennett. Es obvio que la licencia poética le permite al profesor Dennett decir que los termostatos *algo así como creen* que hace demasiado calor y por consiguiente apagan la calefacción central. Pero esto no aporta nada a la ingeniería ni a la explicación de los mecanismos homeoestáticos.

El profesor Dennett afirma (pág. 88) que nosotros no nos ocupamos de sus intentos de utilizar lo que el llama «la actitud intencional» para explicar los procesos corticales. De hecho, hablamos con cierta extensión de su idea de actitud intencional (PFN, págs. 427-431), dando siete razones para dudar de su inteligibilidad. Dado que el profesor Dennett no ha respondido a estas objeciones, no tenemos, de momento, nada más que añadir al tema.

En el debate de la APA, el profesor Dennett declaró que hay «cientos, tal vez miles, de experimentos» que demuestran que una parte del cerebro alberga información que contribuye a «un proceso de interpretación activo en otra parte del cerebro». Esto, insistió, es «algo así como aseverar —algo así como decir "Sí, aquí hay color", "Sí, aquí hay movimiento"». Esto, dijo, «es sencillamente obvio». Pero el hecho de que las células de la corteza estriada visual se exciten en respuesta a los impulsos que se transmiten desde la retina no significa que *interpreten* o *algo así como interpreten* nada. ¿O deberíamos también decir que un infarto demuestra que el corazón tiene algo así como información sobre la ausencia de oxígeno en la corriente sanguínea y algo así como interpreta esto como señal de una obstrucción coronaria? ¿O que la linterna que no se enciende posee in-

formación sobre la cantidad de corriente eléctrica que llega a la bombilla y la interpreta como señal de que sus pilas se han agotado?

24. El pensamiento no ocurre *en* el ser humano, sino que el ser humano lo realiza. El *acontecimiento* de mi pensamiento de que ibas a hacer X se localiza dondequiera que yo estuviese cuando pensaba eso; el *acontecimiento* de que yo veía que estabas haciendo X se localiza dondequiera que yo estuviese cuando te vi haciendo X. Éste es el único sentido en que pensar, percibir, etc., tienen una ubicación. Preguntarse, como hace el profesor Searle (pág. 110), dónde *exactamente* ocurrió el pensamiento, en cualquier *otro* sentido, es como preguntar dónde exactamente una persona pesó 80 kilos en algún sentido distinto del que especifica la respuesta «Cuando estuvo en Nueva York el año pasado». Las sensaciones, por el contrario, tienen una ubicación somática —si me duele la pierna, tengo un dolor en la pierna. Mi estado (si es que se trata de un estado) de tener un dolor en la pierna se dio dondequiera que estuviese cuando me dolía la pierna.

25. Para pensar o andar es necesario que el cerebro funcione con normalidad, pero uno no *anda* con su cerebro. Tampoco piensa *con* él en mayor medida que oye o ve con él.

26. El profesor Searle arguye que, dado que repudiamos los *qualia* tal como los entienden los filósofos, no podemos dar respuesta a la pregunta de en qué consiste el *pasar por un proceso mental* (pág. 111). Si *recitar el alfabeto en la imaginación* (el ejemplo del profesor Searle) cuenta como proceso mental, consiste en que uno empieza por decirse a sí mismo «a», luego «b», luego «c», etc., hasta llegar a «x, y, z». Este proceso mental no se identifica por su sensación cualitativa, sino por el hecho de que es la recitación del alfabeto. Entre los criterios para su ocurrencia está que el sujeto lo diga. Se podría suponer, por supuesto, que va acompañado de procesos neurales hasta ahora desconocidos, cuyo lugar aproximado se puede identificar mediante la correlación inductiva a partir del uso de imágenes por RMf.

27. Descartes, *Principles of Philosphy*, 1:46, 67, y en especial 4:196. (trad. cast.: *Los principios de la filosofía*, Barcelona, RBA, 2002).

28. Como afirma el profesor Churchland en «Cleansing Science», págs. 469 y sigs., 474.

29. *Ibid.*, pág. 470.

30. Para más detalles, véase P. M. S. Hacker, *Wittgenstein: Meaning and Mind*, parte 1: *The Essays* (Blackwell, Oxford, 1993), «Men, Minds and Machines», págs. 72-81.

31. C. Blakemore, «Understaning Images in the Brain», en H. Barlow, C. Blakemore y M. Weston-Smith, (comps.), *Images and Understanding* (Cambridge University Press, 1990), pág. 265. (trad. cast.: *Imagen y conocimiento*, Madrid, Crítica, 1994).

32. J. Z. Young, *Programs of the Brain* (Oxford, Oxford University Press, 1978), pág. 112 (trad.cast.: *Los programas del cerebro*, Madrid, Debate, 1983).

33. D. Chalmers, *The Conscious Mind* (Oxford, Oxford University Press, 1996), pág. 4 (trad. cast.: *La mente consciente*, Barcelona, Gedisa, 1999).

34. F. Crick, *The Astonishing Hypothesis* (Londres, Touchstone, 1995), págs. 9 y sigs. (trad. cast. *La búsqueda científica del alma. Una revolucionaria hipótesis para el siglo XXI*, Madrid, Debate, 1994).

35. A. Damasio, *The Feeling of What Happens* (Londres, Heineman, 1999), pág. 9 (trad. cast.: *La sensación de lo que ocurre*, Madrid, Debate, 2001).

36. Ned Block, «Qualia», en S. Guttenplan, (comp.), *Blackwell Companion to the Philosophy of Mind* (Oxford, Blackwell, 1994), pág. 514.

37. Searle, *Mystery of Consciousness* (Londres: Granta, 1997), pág. xiv (trad. cast.: *El misterio de la conciencia*, Barcelona, Paidós, 2000).

38. T. Nagel, «What Is It Like to Be a Bat?» reimpreso en su *Mortal Questions* (Cambridge, Cambridge University Press, 1979), pág. 170. (trad. cast.: «Qué es ser un murciélago», en *Ensayos sobre la vida humana*, México, Fondo de Cultura Económica, 2000).

39. El profesor Searle (al igual que Grice y Strawson) supone que las experiencias perceptivas se caracterizan desde la perspectiva de su grado

de comunidad con las experiencias ilusorias y alucinatorias. Desde esta perspectiva, toda experiencia perceptiva es, por así decirlo, una alucinación, pero la percepción verídica es una alucinación con un tipo especial de causa. Esto, desde nuestro punto de vista, es un error.

40. El profesor Searle afirma que negamos la existencia de las experiencias cualitativas (pág. 99). Desde luego no negamos que las personas tengan experiencias visuales, por ejemplo, que ven cosas. Tampoco negamos que el ver cosas pueda tener ciertas cualidades. Lo que negamos es que, siempre que alguien ve algo, haya algo que sea como ver lo que ve, y aún menos algo que se siente como ver lo que uno ve. Y negamos que «la sensación *cualitativa* de la experiencia» sea su «esencia definitoria» (pág. 115). El ver o el oír no se definen por referencia a lo que los sujetos sienten, sino por referencia a lo que nos permiten detectar.

41. G. Wolford, M. B. Miller y M. Gazzaniga, «The Left Hemisphere's Role in Hypothesis Formation», *Journal of Neuroscience,* 20 (2000), RC 64 (1-4), pág. 2.

42. Nuestro agradecimiento a Robert Arrington, Hanoch Ben-Yami, Hanjo Glock, John Hyman, Anthony Kenny, Hans Oberdiek, Herman Philipse, Bede Rundle y en especial a David Wiggins, por sus útiles comentarios al primer borrador de este artículo, que presentamos en el debate «Autores y Críticos», que la División del Este de la APA organizó en Nueva York el 28 de diciembre de 2005.

Epílogo

1. Así lo sugirió el profesor Searle en el encuentro de la APA. Afirmó que tal objetivo se alcanzará en tres pasos: primero, determinar los correlatos neurales de la conciencia (CNC); segundo, establecer la relación causal entre la conciencia y estos CNC; y, por último, desarrollar una teoría general que relacione la conciencia y los CNC.

2. J. Searle, *Mind: A Brief Introduction* (Oxford University Press, 2004), cap. 5.

3. J. Searle, «Consciousness: What We Still Don't Know», *New York Review of Books*, 13 de enero de 2005. Se trata de una reseña crítica de la obra de Christof Koch *The Quest for Consciousness* (Greenwood Village, CO, Roberts, 2004).

La búsqueda continúa

1. Un estudio especialmente iluminador sobre esta obra es el de Lina Cable, «Such nothing is terrestriall; philosophy of mind on Phineas Fletcher's Purple Island». *Journal of Historical Science* 19(2), págs. 136-152.

2. Aristóteles, *On the soul*, 403^{a25}-403^{b1}, en Richard McKeon (comp.), *The Basic Works of Aristotle*, trad. J. A. Smith (Nueva York, Random House, 1941) (trad. cast.: *Acerca del alma*, Madrid, Gredos, 1978, pág. 135).

3. *Ibid.*, 408^{b10-15}.

4. En sus *Philosophical Investigations*, §265 (trad. cast.: *Investigaciones filosóficas*, Barcelona, Crítica, 1988).